



QY	701	GGGCTCCAGCTGCACTGTCACGTGGGCGGGGTATGTGGCCCCCTCAGTAGGCTCTGACG	760
Db	499	GGCTTCACCTGCACTGTACTGTGGCTGGGGGTATGTGGCCCCCTCAGTAGGCTCTGACG	558
QY	761	CCCAAGCCACTGCAAGCAACTCGAGGTGCTCTGATCAGTGTGAGACGTGTAACTGCTT	820
Db	559	CCCAAGCCACTGCAAGCAACTCGAGGTGCTCTGATCAGTGTGAGACGT - GTAACGTGCT	617
QY	821	GTAAACAATGACCCCAAGCCCTGAGAGACCCGACATTGTCCAAGAAGACATGTGTGTGC	880
Db	618	GTAAACAATCAACGCGCAAGCCCTGAGAGACCCGACATTGTCCAAGAAGACATGTGTGTGC	677
QY	881	TGGCTATGTGAGAGGGGGGCAAGAGACGCTGCGAGGGTACTCTGGGGGCCACTCTCCTG	940
Db	678	TGGCTATGTGAGAGGGGGGCAAGAGACGCTGCGAGGGTACTCTGGGGGCCACTCTCCTG	737
QY	941	CCCTGTGGAGGGTCTCTGTGTACTGACGGGCAATTGTGACTGGGGAGATGCTGTGGGGC	1000
Db	738	CCCTGTGGAGGGTCTCTGTGTACTGACGGGCAATTGTGACTGGGGAGATGCTGTGGGGC	797
QY	1001	CCGGAACAAGGCTGGTGTGTACACTCTGAGCCCTCAAGTATGCTCTGAGATCCAAAGCA	1060
Db	798	CCGGAACAAGGCTGGTGTGTACACTCTGAGCCCTCAAGTATGCTCTGAGATCCAAAGCA	857
QY	1061	GGTACACAACACTCCAGCCTGTGTGTGTGCCCCCAAAACCAAGAGTCCAGCCGACAGCA	1120
Db	858	GGTACACAACACTCCAGCCTGTGTGTGTGCCCCCAAAACCAAGAGTCCAGCCGACAGCA	917
QY	1121	CCTCTGTGTGGACCACTGTGGCCTTACAGTCT	1151
Db	918	CCTCTGTGTGGACCACTGTGGCCTTACAGTCT	948

```

RESULT 2
US-09-386-642-7
Sequence 7, Application US/09386642
Patent No. 6420157
GENERAL INFORMATION:
APPLICANT: Darrow, Andrew
APPLICANT: Qi, Jensen
APPLICANT: Andrade-Gordon, Patricia
TITLE OF INVENTION: Zymogen Activation System
FILE REFERENCE: OKT-1028
CURRENT APPLICATION NUMBER: US/09/386,642
CURRENT FILING DATE: 1999-08-31
NUMBER OF SEQ ID NOS: 60
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 7
LENGTH: 1169
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: Fusion gene
OTHER INFORMATION: With homo sapien serine protease catalytic domain
US-09-386-642-7

```

Query Match	46.7%	Score 778.2	DB 4	Length 1169
Best Local Similarity	98.3%	Pred. NO. 5.7e-175		
Matches 797; Conservative	0	Mismatches 13	Indels 1	Gaps 1

Qy	341	ATCAAGGATGGCAGCAGTGCAGTCCCGGTCAGTGGCCCTGGCAAGTACATCAACAT	400
Db	166	ATCGTGGGGGGTATGCTCTAGAGGCCGGTCAGTGGCCCTGGCAAGTACATCAACAT	225
Qy	401	GAAGGCGTCACATGTTGTGTGGTCTCTCGTGTCTGACAGATGGGTGCTGTCACTCT	460
Db	226	GAAGGCGTCACATGTGTGTGTGGTCTCTCGTGTCTGACAGATGGGTGCTGTCACTCT	285
Qy	461	CATGTCTTCCCTAGCAGACCAAGAAAGCTTATGAGGTCAAGCTGGGGGGCCACAG	520
Db	286	CATGTCTTCCCTAGCAGACCAAGAAAGCTTATGAGGTCAAGCTGGGGGGCCACAG	345

OY	521	CTGACCTCTCACTCCGAGAGAGCGCCAAAGTATGAGACCTGAAAGGACATCATCCCCACCC	580
Db	346	CTAGACTCTCACTCCGAGAGAGCGCCAAAGTATGAGACCTGAAAGGACATCATCCCCACCC	405
OY	581	AGTCACTCCAGAGAGGAGCTCCAGAGGAGACATTGACTCTCCAACTCAGCAGACCATC	640
Db	406	AGTCACTCCAGAGAGGAGCTCCAGAGGAGACATTGACTCTCCAACTCAGCAGACCATC	465
OY	641	ACCTTCTCCCGCTACATCCGGGCCCATCTGCTCCCTGACGCAACGCGCTCCTCCCAAC	700
Db	466	ACCTTCTCCCGCTACATCCGGGCCCATCTGCTCCCTGACGCAACGCGCTCCTCCCAAC	525
OY	701	GAGCTCCACTGCACTGTCACTGAGTGGAGTCAATGTGAGCCCTTCAGTAGAGCTCTGACG	760
Db	526	GAGCTCCACTGCACTGTCACTGAGTGGAGTCAATGTGAGCCCTTCAGTAGAGCTCTGACG	585
OY	761	CCCAAGCACTGCAAGCACTGAGAGTGCCTCTGATCATGTGCTGAGAGTGTAACTGCT	820
Db	586	CCCAAGCCACTGCAAGCACTGAGAGTGCCTCTGATCATGTGCTGAGAGTGTAACTGCT	644
OY	821	GTAAACATGAGAGGCCAAGCCTGAGAGCGCGCACTTGTCCAAAGAGCAATGAGTGTGC	880
Db	645	GTAAACATGAGAGGCCAAGCCTGAGAGCGCGCACTTGTCCAAAGAGCAATGAGTGTGTGC	704
OY	881	TGAGCTATGTGAGAGGGGGGCAAGAGCGCTCGAAGGTGACTGTGGGGGCCCACTCTCTG	940
Db	705	TGAGCTATGTGAGAGGGGGGCAAGAGCGCTCGAAGGTGACTGTGGGGGCCCACTCTCTG	764
OY	941	CCCTGTGAGAGGCTTCTGTGTAACCTGAAGGGCAATGTATAGCTGGGGGAATGTGCTGAGGCG	1000
Db	765	CCCTGTGAGAGGCTTCTGTGTAACCTGAAGGGCAATGTATAGCTGGGGGAATGTGCTGAGGCG	824
OY	1001	CCGCAACAGGCGTGTGTGTATCACTCTGAGCTCCAGCTATGCTCTCTGATCCAAAGCA	1060
Db	825	CCGCAACAGGCGTGTGTGTATCACTCTGAGCTCCAGCTATGCTCTCTGATCCAAAGCA	884
OY	1061	GGTACAGAACTCCAGGCTGTGTGTGTCGCCCAAAACCAAGAGTCCAGCCGACAGCA	1120
Db	885	GGTACAGAACTCCAGGCTGTGTGTGTCGCCCAAAACCAAGAGTCCAGCCGACAGCA	944
OY	1121	CCCTGTGAGAGCGACCCAGGCTTCAGCTCT	1151
Db	945	CCCTGTGAGAGCGACCCAGGCTTCAGCTCT	975

```

RESULT 3
US-09-387-375-1
: Sequence 1, Application US/09387375
: Patent No. 6485957
: GENERAL INFORMATION:
: APPLICANT: Darrow, Andrew
: APPLICANT: Andrade-Gordon, Patricia
: APPLICANT: Qi, Jensen
: TITLE OF INVENTION: DNA Encoding the Human Serine
: TITLE OF INVENTION: Picotase EOS
: FILE REFERENCE: ORT-1031
: CURRENT APPLICATION NUMBER: US/09/387,375
: CURRENT FILING DATE: 1999-08-31
: NUMBER OF SEQ. ID NOS: 9
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 1
: LENGTH: 1613
: TYPE: DNA
: ORGANISM: Homo sapiens
US-09-387-375-1

```

Query Match	12.4%	Score 207.2	DB 4	Length 1613
Best Local Similarity	57.4%	Pred. No. 6.8e-40		
Best Local 413, Conservative	0	Mismatches 303	Indels 4	Gaps 2

QY	338	CGATCACAGGTGCGACAGTCAGTCGCGCGGTTCAGTGGCCCTGGCAGGTACAGATCAC	397
Db	174	CGGATCGTTGGGGCCCGGATGGCCGGACCGGAGAGTGGCCCTGGCAGGCGAGATCCAG	233

OY		398	TATAGAAAGCGTCAATGTGTGAGTAGAGCTCTCTCCGGTCTGACGAATGGAGTGCTGCACACT	457
Db		234	CATCTGGGGGACAAGTGTGCGGGGGGTTCGTCAATGCCCCCCAATGGAGTCTGAAGAAGC	293
OY		458	GCTCACTGCTTTCCCAGCGACACCACAAGAAAGCTTAATGAGTCAAGCTGGGGGCTCAC	517
Db		294	GCGCATGCTTTCCCCAGAGAGGSCACTGCGACGTGAGTACCCGATGCGCTGGGGGACGTG	353
OY		518	CAGCTAGACTCTACTCCGAGAGAGCGCAAAGTCAACGACACCTGGAAGACATCATCCCCAC	577
Db		354	GCTTGGGCTTCACTTCGCCCCGACGCTCTCGGTGCCCCGTGCGACGGAGTGTCTTGCCC	413
OY		578	CCGAGCTAACCTCCAGAGAGGCTCCGAGGGCGCATTTGACACTCTTCCAATCAGCAGACCC	637
Db		414	CCGGAATACTCCAGAGACGGGGGCCCGGCGACCTGAGCATGCTCAGACTGCTGTGCACCG	473
OY		638	AATACCTTTCTCCGCTAATATCCGAGCCCATCTTGCCTCCTCGACGCCAAAGCCTCTTCCCC	697
Db		474	GTGCCCCCTGAGCGCTCGGTCTCAACCGGTCTTGCCTCGCCCGTGC CGGCGCCCGCCGCG	533
OY		698	AACGAGCTCCACTGCACATGTCACTGATGCGAGGGTCAATGAGCCGCCCTCAGTAGACCTCTG	757
Db		534	CCCGGACACCATGCGGGGTACACGGGCTGCGGGGACGCTCCGCCAGAGATGCTCCCTCCA	593
OY		758	ACGCCCCAAGCCACTGACGACAACTGAGGTGCTCTGATCACTCGTAGACGTGTAACTG	817
Db		594	GAGTGGCGACCGCTACAAAGAGATGAAGGGTGCCTGCT - GGAATTCGCGCACCTTGCGAG	652
OY		818	CCCTTACAAATATGACGCG --- CAAAGCCGAGAGAGAGCCGACATTGTGCCAAAGAGAAATGT	874
Db		653	CTCTTACCACTGGAGCGCGAGACGTGCCCCAGGCTGAGGCAATTGTGCTCTGGAGTCT	712
OY		875	GTGTGCTGTGATGTGAGAGGGGCGAAGAGACGCTGCCAGGATGATCTGTGGGGGCCACT	934
Db		713	GTGTGCGGGCTAACCCGCCAGGGCCACAAGAGACGCTGCGCAGAGGTGATTTCTGGGGGACCT	772
OY		935	CTCGTGGCCGTGTGAGAGGGTCTCTGTGATCTGACGGGCAATTGTGAGCTGGGGAGATGCTG	994
Db		773	GACCTGCTTGCAGTCTGTGGAGCTGGGTCTGTGTGGGCGTGTGTGAGCTGGGGCAAGGCTTG	832
OY		995	TGGGGCCCGCAACAGGCTGTGTGTATCACTCTGGCCCTCCAGCTATGCTCTCTGGATCCA	1054
Db		833	TGCCCTGCGCAACCGTCCAGGGGTTCTAACCAAGTGTGGCCCAATATTAACCCCTGATITCA	892
<hr/>				
RESULT 4				
US-09-387-375-8				
:	Sequence 8, Application US/09387375			
:	Patent No. 6485957			
:	GENERAL INFORMATION:			
:	APPLICANT: Darrow, Andrew			
:	APPLICANT: Andrade-Gordon, Patricia			
:	TITLE OF INVENTION: DNA Encoding the Human Serine			
:	FILE REFERENCE: ORT-1031			
:	CURRENT APPLICATION NUMBER: US/09/387,375			
:	NUMBER OF SEQ ID NOS: 9			
:	SOFTWARE: PatentIn Ver. 2.0			
:	SEQ ID NO 8			
:	LENGTH: 1130			
:	TYPE: DNA			
:	ORGANISM: Artificial Sequence			
:	FEATURES:			
:	OTHER INFORMATION: Description of Artificial Sequence: Nucleic acid			
:	OTHER INFORMATION: sequence of EOS zymogen fusion gene			
:	US-09-387-375-8			

Query Match	12.3%	Score 205.8	DB 4	Length 1130
Best Local Similarly	57.3%	Pred. No. 1.3e-39		
Matches 411	Conservative	0	Mismatches 302	Indels 4
				Gaps 2

QY	341	ATCACAGGAGGAGAGAGTGCAGTATGCGCGGTATGATGAGCCCTGAGGACGGTACAGATACCTAT	4000
Db	166	ATGTGTTGGGGGCTATGCTCTTAGAGACGAGAGTGGCTTGGACAGGCGAGATCAGACT	2250
QY	401	GAAGCGTCCATATGTGTGTGTGGCTCTCGTGTCTGAGCAGTGGAGTGTCTGACTGCT	4600
Db	226	CCCGGGGGACACAGTGTGCGGGGGGTGCTCATTCGCCCCCAGTGGGTGTACACGCGG	2850
QY	461	CACGTCTTCCCGAGCGAGCACCAAGAGAGCTTAGAGTCAAGCTTGGGGGCCACAG	5200
Db	286	CACGTCTTCCCGAGAGGGACATGCGACACTGATAGTACGCGTGGCCTTGGGGGCGCTCGT	3450
QY	521	CTTAGACTCTTACTCCGAGGAGCGCCAAAGGTACAGCACCTGAAAGACATATCTCCCAACCC	5800
Db	346	CTGGGGCTTCACTTCGCCCGCACGCTCTCGGTGCCCCGTGCGACGAGGTGTCTGTCCCCCG	4050
QY	581	AGCTACCTCCAGAGAGGGCTCCAGGGGCGACATTTGACTCTCCCACTCAGACAGCCATC	6400
Db	406	GACTTACTCCGAGAGACGGGGCCCGGGGACACTGGGACTGCTGTGACGTGTCTGCCCCGTG	4650
QY	641	ACCTTCTCCCGCTACATCGAGCCCATCTGCTCTCTGACAGCAACGCTCTTCCCCAC	7000
Db	466	CCCTTAGGGCTCTGGGTCCAACTCGTCTGCTGTGCGCGGCGCCGCGCGCGCGCC	5250
QY	701	GGCTTCCACTGCACTGTCACTGGCTTGGGGTCAATGTGCCCCCTCAGTAGAGCTCTTGACG	7600
Db	526	GGCACACATGCGCGGGTACCGGCTTGGGGGACGCTCGGCCAGAGATGTCCCTCCAGAG	5850
QY	761	CCCAAGCCACTGACAGCACTGAGGTGCTCTGATCAGTCTGTGAGAGCTGTATCTGCTT	8200
Db	586	TGGCGACCGCTTCAAGAGAGTAAAGGTGTGCGGTGTCT-GGACTCGGCGACCTGCGACGGGCTT	6440
QY	821	GTTACAATCTGACGC---CAAGCTGAGAGACCCGCACTTTGTCCAAAGAGCATGTGTGTG	8770
Db	645	CTTACACGTGGGCGGGAGATGCCCCAGGCTGAGCGCATTTGTGTGCTGTGGAGTCTGTG	7040
QY	878	TGCTGGCTATGTGAGAGGGGGCAAGAGCGCTGCGAGGGTACTCTTGGGGGCCACTCTC	9370
Db	705	TGCGGGCTACCCCGAGGGCACCAAGAGCGCTGCGACAGGGTATTTCTGGGGGACCTTGAC	7640
QY	938	CTGCGCTGTGAGAGGTCTTGTGTACTTACGCGGCATTGTGAGCTTGGGGAAGATGCTGTGG	9970
Db	765	CTGCTCTGACGTCTGGAGAGCTGGGTCTGTGTGGTGTGGTGTGAGCTGGGGGCAAGGGTGTGTG	8240
QY	998	GGCGCGCAACAGGCTGTGTGTATACCTTGGCTCCAGCATATGCTCTGATATCCA	1054
Db	825	CTGCGCAACCGTCCAGGGGTCTTACACAGTGTGGCCACATATAGCCCTCTGGATTTCA	861

```

RESULT 5
US-09-386-653A-8
: Sequence 8, Application US/09386653A
: Patent No. 6458554.
: GENERAL INFORMATION:
: APPLICANT: Andrade-Gordon, Patricia
: APPLICANT: Darrow, Andrew
: APPLICANT: Qi, Jian-shen
: TITLE OF INVENTION: DNA encoding the novel human serine
: TITLE OF INVENTION: Protease T
: FILE REFERENCE: ORT-1032
: CURRENT APPLICATION NUMBER: US/09/386,653A
: CURRENT FILING DATE: 1999-08-31
: NUMBER OF SEQ ID NOS: 11
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 8
: LENGTH: 1130
: TYPE: DNA
: ORGANISM: Artificial Sequence
: FEATURE:
: OTHER INFORMATION: Description of Artificial Sequence: Fusion gene of
: OTHER INFORMATION: Protease T in a zymogen activation vector
US-09-386-653A-8

```





Db 536 CCAACCCAAACCATCAAGATGACATGCTGTGGCCGCGCTTGAGAGGGCAAGAGGA 595  
Qy 905 CGCTTCACAGGGTGAAGTCTGGGGGCCCACTCTCTCCGCTGTGGAGGGTCTGTGACT 964  
Db 596 TGCCTGCAAGGGGGAAGTCTGGGGGCCCGCCCTGTGTGTCTGTGTGTGTGTGTGTGT 655  
Qy 965 GACGGGCAATTTGAGTGGGAGATGCTGTGGGGGCCCGCAAGAGGCTGTGTGTGTGTGTGT 1024  
Db 656 GGGGGGGGTGATCAGCTGGGGGTGAGGGGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 715  
Qy 1025 TGTGGCTTCAGCTATGCTCTCTGTGATCCAAAGCAAGTGAAGTCTCAG 1076  
Db 716 CGGTGTACCGCCGCCACCACTGATTCATTCGATTCATCCCAACTGTGAG 767

RESULT 7  
US-09-386-653A-1

/ Sequence 1, Application US/09386653A  
/ Patent No. 6458564  
/ GENERAL INFORMATION:  
/ APPLICANT: Andrade-Gordon, Patricia  
/ APPLICANT: Darrow, Andrew  
/ APPLICANT: Qi, Jian-shen  
/ TITLE OF INVENTION: DNA encoding the novel human serine  
/ TITLE OF INVENTION: protease T  
/ FILE REFERENCE: ORT-1032  
/ CURRENT APPLICATION NUMBER: US/09/386,653A  
/ CURRENT FILING DATE: 1999-08-31  
/ NUMBER OF SEQ ID NOS: 11  
/ SOFTWARE: Patentin Ver. 2.0  
/ SEQ ID NO 1  
/ LENGTH: 1110  
/ TYPE: DNA  
/ ORGANISM: Homo sapiens  
US-09-386-653A-1

Query Match 11.6%; Score 192.8; DB 4; Length 1110;  
Best Local Similarity 56.3%; Pred. No. 1.5e-36;  
Matches 401; Conservative 0; Mismatches 307; Indels 4; Gaps 2;

Qy 368 GGTCAATGGCCCTGGAGAGTCAAGTCACTAATGAAGGGCTCATGTGTGTGTGTGTGTGT 427  
Db 165 GGGAGATGGCCCTGGAGAGTCAAGTCACTAATGAAGGGCTCATGTGTGTGTGTGTGTGT 224  
Qy 428 CTCGTGTGTGAGAGTGGGGT 487  
Db 225 CTCATGCGGAGAGTGGGGT 284  
Qy 488 GAAGCTATGAGGTCAAGTGGGGGCCCAAGCTAGACTCTTAATCCGAGGAGCGCCAG 547  
Db 285 TCCCTGTACAGGTCTGTGGGGGCCAAGGAGCTAGTGCAGCGGGAGCCACACGCTATG 344  
Qy 548 GTACAGACCTTGAAGAGATCATCCCCACCCAGTACTCTCCAGAGAGGCTTCCAGGGC 607  
Db 345 TATGCCCGGGGTGAGGAGTGAAGCAACCCCTGTACAGAGGCAAGGCTTCCAGGGC 404  
Qy 608 GACATTTGCACTCTCCCACTCAAGAGAGCCATCACTTCCCGGCTCAATCCGGGCCCTATC 667  
Db 405 GACGTGGCCCTGTGTGAGTGAAGGACCAAGTCCCTTACCAATTAATCTTCCCCGTG 464  
Qy 668 TGCCTTCCTGAGCAAGCCCTCTTCCCAAGGCTTCCAGTCACTGTCACTGTGTGTGTGT 727  
Db 465 TGCCTTCCTGAGCAAGCCCTCTTCCCAAGGCTTCCAGTCACTGTGTGTGTGTGTGTGT 524  
Qy 728 GGTCAATGGCCCTGTGTGAGTGAAGTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 787  
Db 525 GGCAGGCCAGTGAAGAACTCTGCGCCGAACCGGATCCTGTGAAGAACTCTGCTGTG 584  
Qy 788 CCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 844  
Db 585 CCAATCATGACACACCAAGTGAAGTGTG-CTTTCAGAGCAAGAAAGAGTGTGTGTGTGT 643  
Qy 845 GAGGCGGACCTTTGTTCAGAGAGATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 904

Db 644 CCAACCCAAACCATCAAGATGACATGCTGTGGCCGCGCTTGAGAGGGCAAGAGGA 703  
Qy 905 CGCTTCACAGGGTGAAGTCTGGGGGCCCACTCTCTCCGCTGTGGAGGGTCTGTGACT 964  
Db 704 TGCCTGCAAGGGGGAAGTCTGGGGGCCCGCCCTGTGTGTCTGTGTGTGTGTGTGTGTGT 763  
Qy 965 GACGGGCAATTTGAGTGGGAGATGCTGTGGGGGCCCGCAAGAGGCTGTGTGTGTGTGTGT 1024  
Db 764 GGGGGGGGTGATCAGCTGGGGGTGAGGGGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 823  
Qy 1025 TGTGGCTTCAGCTATGCTCTCTGTGATCCAAAGCAAGTGAAGTCTCAG 1076  
Db 824 CGGTGTACCGCCGCCACCACTGATTCATTCGATTCATCCCAACTGTGAG 875

RESULT 8  
US-09-620-312D-431

/ Sequence 431, Application US/09620312D  
/ Patent No. 6569662  
/ GENERAL INFORMATION:  
/ APPLICANT: Tang, X. Tom  
/ APPLICANT: Liu, Chenghua  
/ APPLICANT: Asundi, Vinod  
/ APPLICANT: Zhang, Jie  
/ APPLICANT: Ren, Feiyan  
/ APPLICANT: Chen, Rui-hong  
/ APPLICANT: Zhao, Qing A.  
/ APPLICANT: Wehrman, Tom  
/ APPLICANT: Xue, Aidong J.  
/ APPLICANT: Yang, Yonghong  
/ APPLICANT: Wang, Jian-Rui  
/ APPLICANT: Zhou, Ping  
/ APPLICANT: Ma, Yunding  
/ APPLICANT: Wang, Dunrui  
/ APPLICANT: Wang, Zhiwei  
/ APPLICANT: John, Tillinghast  
/ APPLICANT: Drmanac, Radoje T.  
/ TITLE OF INVENTION: No. 6569662el Nucleic Acids and  
/ TITLE OF INVENTION: Polypeptides  
/ FILE REFERENCE: 784CIP2B  
/ CURRENT APPLICATION NUMBER: US/09/620,312D  
/ PRIOR FILING DATE: 2000-07-19  
/ PRIOR APPLICATION NUMBER: 09/552,317  
/ PRIOR FILING DATE: 2000-04-25  
/ PRIOR APPLICATION NUMBER: 09/488,725  
/ NUMBER OF SEQ ID NOS: 1105  
/ SOFTWARE: pt\_FL\_genes Version 1.0  
/ SEQ ID NO 431  
/ LENGTH: 1212  
/ TYPE: DNA  
/ ORGANISM: Homo sapiens  
/ FEATURE:  
/ NAME/KEY: CDS  
/ LOCATION: (135)..(1007)  
US-09-620-312D-431

Query Match 11.5%; Score 191.2; DB 4; Length 1212;  
Best Local Similarity 56.2%; Pred. No. 3.8e-36;  
Matches 400; Conservative 0; Mismatches 308; Indels 4; Gaps 2;

Qy 368 GGTCAATGGCCCTGGAGAGTCAAGTCACTAATGAAGGGCTCATGTGTGTGTGTGTGTGT 427  
Db 264 GGGAGTGGCCCTGGAGAGTCAAGTCACTAATGAAGGGCTCATGTGTGTGTGTGTGTGT 323  
Qy 428 CTCGTGTGTGAGAGTGGGGT 487  
Db 324 CTCATGCGGAGAGTGGGGT 383  
Qy 488 GAAGCTATGAGTCAAGTGGGGGCCCAAGCTAGACTCTTAATCCGAGAGCGCCAG 547  
Db 384 TCCCTGTACAGGTCTCTGTGGGGGCAAGGAGCTAGTGAAGCGGAGACACAGCTATG 443

QY	548	GTCAAGACCCGGAAGGATATATCCCAACCCCACTCATCTCAAGAGGGCTCCAGGGC	607
Db	444	TATGCGCGGGTGAAGGACAGGTGAGAGCAACCCCTGTATCAAGGGACAGGCTTCACGCTT	503
QY	608	GACATTTGACTCTTCCAACTCAGACAGACCCATCACCTTCTCCGGTACATTCGGGCCATC	667
Db	504	GACGTGGCCCTGTGTGGAGCTGGAAGGACAACAGTCCCTTACCAATTACATCTCTCCCCGTG	563
QY	668	TGCCTCCCTGACAGCCACGCGCTCTTCCCAAAGGCTTCCATGTGCACTGTCACTGGCTGG	727
Db	564	TGCCGTGCTGACCCCTCGGTGATCTTTTGAGACGGGCATGAATGTGCTGGGCTTCATGCTGG	623
QY	728	GGTATGTGGCCCCCTCATGTAGCTCTCTGACGCCCAAGCCACTGACAGATTCGAGGTG	787
Db	624	GGCAGCCCCAGTAGAGGAAGACCTCTGCGCCGAACCGCGGATCTCTCGAAGAACTCGTGTG	683
QY	788	CCTGTGATCAGTCGTTGAGACGTTGTTAACTGCTGTATCAACATTCGACGCCAAGCTGTG--A	844
Db	684	CCCATCATTCGACACACCCAGTGCACCTTG-CTTTACAGCAAGACACCGAGTTTGGCTA	742
QY	845	GGAGCCGCACTTTGTTCAGAGAGAGCATGGTGTGTGCTGTGCTATGTGAGGGGGGCAAGA	904
Db	743	CCAAOCCAAAACCATCAAGAAATGACATGTGTGTGGCGGCTTCAGAGAGGGCAAGAGGA	802
QY	905	CGCTGCGAAGGTGACTCTTGGGGGGCCCACTCTCTGCGCTGTGTGAGAGGGTCTTGTA	964
Db	803	TGCCTGCAAGGGCGACTCGGGGGGGCCCCCTGTGTGTGCTCTGTGGGTCAATGCTGGCTGCA	862
QY	965	GACGGGCAATTGTGACCTGGGGAGATGTCTGTGGGGGCCCGCCACAGGCTGTGTGTGAC	1022
Db	863	GGCGGGGGTGAATCACTGGGGGTGAGGGGCTGTCCGCCAGAACCGCCAGGTGTCTCAAT	922
QY	1025	TCTGGCTTCCAGCTATGCTCTCTGGATTCCAAGACAGGTGACAGAACTCCAG	1076
Db	923	CCGTTCACCGCCACCAACACTGGATTCATGTGATCATCTCCCAATCTCAG	974

```

RESULT 9
US-09-386-629-2
Sequence 2, Application US/09386629
Patent No. 6426199
GENERAL INFORMATION:
APPLICANT: Darrow, Andrew L.
APPLICANT: Qi, Jensen
TITLE OF INVENTION: Identification and Characterization of the complementary
FILE REFERENCE: ORT-1030
CURRENT APPLICATION NUMBER: US/09/386,629
CURRENT FILING DATE: 1999-08-31
NUMBER OF SEQ ID NOS: 11
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 2
LENGTH: 1166
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: C-E catalytic
US-09-386-629-2

```

Query Match	9.2%	Score 153	DB 4	Length 1166
Best Local Similarity	53.3%	Pred. No. 4e-27		
Matches 400	Conservative	0	Mismatches 335	Indels 16
			Gaps	3
OY	341 ATCAGAGTGGCAGCAGAGTGCAGTCCCGGTCTAGTGCCTGTGCAGAGTGCAGATCAGATCCTAT	400		
Db	166 ATCGTTGGGGGTTATGCTCTTAAGAGACACGAGTGGCCTGGATGTGTGACATTCAGGAAG	225		
OY	401 GAAGGCGTTCATATGTGTGTGTGGCTCTCTCGTGTGACAGATGGGTGTGCAGTGTCT	460		
Db	226 AATGGAGACCCACACTGCGCAGAGTTCTTGTCTCACCAGCCGCTGGGTATATCTGTGTGCC	285		

OY		461	CACGCTTCCCGACGAGCACCAAGAAGCC---TATAGGTTCAGACTGGGGCCAC	517
Dd		286	CACGTITTTCAAAGACAACCTGAACAACCATACTCTTTCTCTGTGCTGTGGGGCCTGG	345
OY		518	CAAGTAG--ACTCTTACTCCGAGGAGCCCAAGGTCAAGACCCTGAAGACATATCCCC	574
Dd		346	CAGCTGGGAAACCTTGAGCTCTGGTCTCCAGAAGGTGGGTGTGTGCTGGGTGAGACCCAC	405
OY		575	CACCCCAAGTACCTCCAGGAGGGGTCCAGGGGGGACATTGTGCATCTTCCMACTGACAGA	634
Dd		406	CTGTGTATTCTTGAAGAAAGTGGCTGTGACAGACTTGGCTTGTGTGCTGTGAAGGC	465
OY		635	CCCATCACTTCTCCCGCTACATCCGGGCCCATCTGCCTCCCTGAGCCMAGCCTCTTC	694
Dd		466	TCCATAGAGTTCTCAGAGGGGGTCTGGCCATCTGCCATCTGACTGATGCTTATCAACTTC	525
OY		695	CCCAACGGCTCCACTGTCGACTGTTCATCTGTGGGGTCAATGTGGCCCTCAGTAGACTC	754
Dd		526	CTTCCAACACCCCACTCTGTGATTTCAAGGCTGGGGGAGACATCCAAGATGGAGTTCCCTTG	585
OY		755	CTGACGGCCMAAGCCACTGCAGCAACTCGAAGGTGCTCTGATCAATCTGAGACGTGTAA	814
Dd		586	CCCCACCTTCAGACTCTGCAAGACTGAAGTTCTTATCATGACTGGAAGTCTG----	641
OY		815	CTGCTGTACAACATCGACGCCCAAGCCTGAAGACCAGCATTTTGTCCAAAGACATGCT	874
Dd		642	-----CAGCCATCTGTACTGTGGCGGGAGCAGACAGGGAACCCATCACTGAGGACATCT	695
OY		875	GTTGTGCTGTATATGTGAAGGGGGSCAAGANCGCTTCAGGGGTACATCTGGGGGCCACT	934
Dd		696	GTGTGCGGGCTACTTGAAGGGGGAGCGGAATGTTGTGTGGGCACATCTCGGGGGCCCTT	755
OY		935	CTTCTGCCCTTGTGAGGGGTCTCTGTACTCTGACCGGCAATTGTGACTGTGGAGAGATGCTGTG	994
Dd		756	CATGTGCAAGTGAACGGCGCTTGT	815
OY		995	TGGGGCCCGCAACAGGCTGTGTGTATCACTTGTGACTTCAGCTATGTGCTTCTTGATTTCA	1055
Dd		816	TGCCGAGGCAACAGGCCCGGGGTTCTAATCAAGCCTCTGTGCGACCGCTCTGTGGTGA	875
OY		1055	AAGCAAGTGCACAGACTCCAGCCTGTGTG	1085
Dd		876	GAAATGTGTCAAGGGGTGCAAGTCTCCGGGS	906

RESULT 10  
 US-09-023-942A-28  
 Sequence 28, Application US/09023942A  
 Patent No. 6479274  
 GENERAL INFORMATION:  
 APPLICANT: (US only) ANTALIS Toni Marie and HOOVER John David  
 TITLE OF INVENTION: NOVEL MOLECULES  
 NUMBER OF SEQUENCES: 30  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: SCULLY, SCOTT, MURPHY & PRESSER  
 STREET: 400 GARDEN CITY PLAZA  
 CITY: GARDEN CITY  
 STATE: NEW YORK  
 COUNTRY: USA  
 ZIP: 11530  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patent In Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/09/023.942A  
 FILING DATE: 13-FEB-1998  
 CLASSIFICATION: 435  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: P05101/97  
 FILING DATE: 13-FEB-1997



```

Db      874  GCTGCTGGCCGGCATCATCACTGGGGGAGAGGCTGTGCCAGCCCAAGGCCCGGGGT  933
QY      1019  GTACACTCTGGCCCTTCAGACTATGCTCTGTGATCCAAAGCAAGGTGACAGAACTCAAGCC  1078
Db      934  CTACATCAGCCCTCTCTGCGCACCGGCTCTGGGTGTGAGAAAGATCTGCAAGGGGTGCACT  993
QY      1079  TCGTGTG 1085
Db      994  CCGCGGG 1000

RESULT 12
US-09-016-366A-18
; Sequence 18, Application US/09016366A
; Patent No. 5955431
; GENERAL INFORMATION:
; APPLICANT: Stevens, Richard L.
; APPLICANT: Huang, Chifu
; TITLE OF INVENTION: MAST CELL PROTEASE PEPTIDE
; TITLE OF INVENTION: INHIBITORS
; NUMBER OF SEQUENCES: 65
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wolf, Greenfield & Sacks, P.C.
; STREET: 600 Atlantic Avenue
; CITY: Boston
; STATE: MA
; COUNTRY: U.S.A.
; ZIP: 02210-2211
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/016,366A
; FILING DATE: January 30, 1998
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/037,090
; FILING DATE: 05-FEB-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Plumer, Elizabeth R.
; REGISTRATION NUMBER: 36,637
; REFERENCE/DOCKET NUMBER: B0801/7093
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-720-3500
; TELEFAX: 617-720-2441
; TELEX:
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1137 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; US-09-016-366A-18

Query Match      8.7%; Score 145.2; DB 2; Length 1137;
Best Local Similarity 53.0%; Pred. No. 2.8e-25;
Matches 439; Conservative 1; Mismatches 359; Indels 29; Gaps 5

QY      257  ATTCTGCTCTATCTTGATTACTCCGGTGGGAGACAGAGCGGAAGGAGCAGAAAGCTTCC  316
Db      3    AATTCGTGCTGTGCTGGCGCTGCGCCGCTGCGGAGAGCGCGCTTACGCGGCCCTGCCCCA  62
QY      317  TCGCGTGTGCCCCCAAGACAGCATCACAGGTGCGACAGATGCAATGCCCGTCAGTGG  376
Db      63  GCGCAGGCGCTTCAGCAGAGTGGGCAATCGTCGGGGGTCAAGAGAGCCCCCAGAGCAAGTGG  122
QY      317  CCTGAGCAGGTGAGCATCACTATGAAGC-----GTCCATGTGTTGGTGGCTCT  427
QY      377  CCTGAGCAGGTGAGCTTGAAGTCCACGAGCCCATCTGATGATGATCTTTCGCGGGGCTCC  182
Db      123  CCTGAGCAGGTGAGCTTGAAGTCCACGAGCCCATCTGATGATGATCTTTCGCGGGGCTCC  182

```

Oy	428	CTGCTGTCTAGAGCATGGTGGTGTCTGTACGCTGCATCGTCTC---CCAGGAGACACAC	484
Db	183	CTATTCACACCCCATGGTGGTGTGAACCGCAGCCGACTGTGGGAGCCGAGCTCAAGAT	242
Oy	485	AAGGAAGCTTATGATGATCTGACCTGGGGGCCACAGCTAGACTCTTCTTCGAGGACGCC	544
Db	243	CTGGCCGCTCTAGAGGTGTGCAACTGCGGGAGCAGCACTTAATACTACAGGACCAACTGTG	302
Oy	545	AAGGTACGCAACCTGAAAGACATCATTCGCCACCCAGCTAAGCTTCCAGAGAGGCTCCAG	604
Db	303	CCGGTTCAGCAGG-----ATCATGTGTGACCCCAAGTTCTACACGCCAGATGGA	353
Oy	605	GGGACATTTGACATCTCTCCAACTCAGCAGACCCATCACTTCTCCGCTACATCGGCCC	664
Db	354	GCGGACATTCGCTCTGTGAGCTGTGAGGAGCCGCTGAACCTTCCAGCCAGCTCACAG	413
Oy	665	ATCTGCTCTCCCTGAGCCAAAGCCTCTTCCCAACGGCCTTCACATGCACTGTCACTGGC	724
Db	414	GTCAACCTTGCCCCCTGTGCTCAGAGACCTTCCCCCGGGATGCGGTGAGTCACTGGC	473
Oy	725	TGGGTGATGTGGCCCCCTCAGTGAAGCTTCTGTAGCGCCAAAGCCATGCGAGCACTGGAG	784
Db	474	TGGGGCATGTGGACAATGATGAGCGCTCCACCGSCATTTCTCTGAAACAGCTGAAG	533
Oy	785	GTGCTCTGTGATCGTGAAGCCTGTGTAACCTGCTGTACAAACATGCAAGCCGAAGCTGA	844
Db	534	GTCCCATATGAAAAACACATTTGTGAGCAAAATACACACTTTGGCGCTTACCGGA	593
Oy	845	GGAGCCG--CACTTGTCCAAAGAGACATGTGTGTGCTGTATGTGAGGGGGCAG	902
Db	594	GACAGCTGCGCATGCTGCGGTGAGCAGATGCTGTGCCGG-----GACACCCGGAGG	647
Oy	903	GAGCCCTGCAAGGGGTGACTGTGGGGGCCACTCTCTGCCCCCTGTGGAAGGCTCGGTAC	962
Db	648	GACTCATGCAAGGGCGACTCGGAGGGCCCCGTGTGTGCAAGGTGAATGCACTGGCTG	707
Oy	963	CTGACGGGCAATTGAGCTGGGGAGATGCCGTGTGGGCCCGCMAAGGCTGTGTGTAC	1022
Db	708	CAGGCGGGCGTGTCACTGTGGGGGAGAGGGCTGTGCCAGGCCMAAGCGCTGGCATTTAC	767
Oy	1023	ACTGTGCGCTTCAGACTATGCTCTCTGTGATCCAAAGCAAGGTGAGAGAA	1070
Db	768	ACCGGTGTCACTACTTGTGACTGTGATTCACCACTATGTGCCCAA	815
RESULT 13			
US-08-978-404B-13			
; Sequence 13, Application US/08978404B			
; Patent No. 5968782			
; GENERAL INFORMATION:			
; APPLICANT: Stevens, Richard L.			
; TITLE OF INVENTION: MAST CELL PROTEASE THAT CLEAVES			
; TITLE OF INVENTION: FIBRINOGEN			
; NUMBER OF SEQUENCES: 74			
; CORRESPONDENCE ADDRESS:			
; ADDRESSEE: Wolf, Greenfield & Sacks, P.C.			
; STREET: 600 Atlantic Avenue			
; CITY: Boston			
; STATE: MA			
; COUNTRY: U.S.A.			
; ZIP: 02210-2211			
; COMPUTER READABLE FORM:			
; MEDIUM TYPE: Diskette			
; COMPUTER: IBM Compatible			
; OPERATING SYSTEM: DOS			
; SOFTWARE: PASCSEQ for Windows Version 2.0			
; CURRENT APPLICATION DATA:			
; APPLICATION NUMBER: US/08/978,404B			
; FILING DATE: 25-NOV-97			
; CLASSIFICATION: 435			
; PRIOR APPLICATION DATA:			
; APPLICATION NUMBER: 60/032,354			



```

Db      245 CTGGCCGCTTGAAGGTGCACTGCGGAGACAGCACTCTTACTTACGAGACCAAGCTGCTG 304
QY      545 AAGGTGAGACCCCTGAAGGACATCATCCCAACCCAGACTCTCCAGAGAGGCTCCAG 604
Db      305 CCGGTGAGAGG-----ATCATGCTGACCCACAGTTCTTACACCCCGGAGATGGA 355
QY      605 GGGCAGATTGACATCTCTCCAACTCAGACAGACCCATCACTTCTCCGCTTACATCCGAGCC 664
Db      356 GGGGACATGCGCTCTGAGAGCTGAGAGAGCCGATGAGGTCTCCAGCCACGTCCACAG 415
QY      665 ATTGCTCTCTCTGAGCAGCAAGCTCTTCTCCCAAGGCTCTCCAGTCACTGCTGCTG 724
Db      416 GTCACTCTGCTCTCTGCTTCAAGACCTTCCCGGAGATGCTGCTGCTGCTGCTGCTG 475
QY      725 TGGGGTCAATGAGCCCTTCAAGAGCTCTGAGCGCCCAAGCCAGCTGAGCAACTGAG 784
Db      476 TGGGGGAGATGAGCAATGATGAGCGCTCCACCGCATTTCTCTGAAAGCAGTGAAG 535
QY      785 GTGCTCTGATCACTGCTGAGACGATGTAACCTGCTGTACATGACGACCAAGCTGA 844
Db      536 GTCCCATATAGAAAAACCACTTTGTGACGCAAAATACGACCTTGGCGCTTACAGGGA 595
QY      845 GAGAGCG--CACTTTGTCAAAGAGACATGCTGTGTGCTGCTATGCTGAGGGGGGCAAG 902
Db      596 GACGACGTCCGATCGCTGCTGACGACATGCTGTGTGCTGCTGCTGCTGCTGCTGCTG 649
QY      903 GAGGCTTGCAGAGGTGATCTGTGGGGGCGCACTGCTGCTGCTGCTGCTGCTGCTGCTG 962
Db      650 GACTCATGCGAGGACGCTCGGAGGGCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 709
QY      963 CTGACGGGCACTTGTGAGCTGAGGAGATGCTGTGGGGCCGCAACAGGCTGTGTGTAC 1022
Db      710 CAGGCGGGGCTGTGAGCTGAGGGGAGAGGCTGTGCGCCAGCCGAGCGGCTGGGATCTAC 769
QY      1023 ACTCTGCTCTCCAGCTTGTGCTCTCTGATCCAAAGCAAGGTGACAGAA 1070
Db      770 ACCGCTGCTCACTACTTGTGATGATCCACCATATGTCTCCCAA 817

```

```

RESULT 15
US-08-978-404B-15
; Sequence 15, Application US/08978404B
; Patent No. 5968782
; GENERAL INFORMATION:
; APPLICANT: Stevens, Richard L.
; TITLE OF INVENTION: MAST CELL PROTEASE THAT CLEAVES
; TITLE OF INVENTION: FIBRINOGEN
; NUMBER OF SEQUENCES: 74
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wolf, Greenfield & Sacks, P.C.
; STREET: 600 Atlantic Avenue
; CITY: Boston
; STATE: MA
; COUNTRY: U.S.A.
; ZIP: 02210-2211
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/978,404B
; FILING DATE: 25-NOV-97
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/032,354
; FILING DATE: 04-DEC-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Plumer, Elizabeth R.
; REGISTRATION NUMBER: 36,637
; REFERENCE/DOCKET NUMBER: B0801/7090
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-720-3500

```

```

; TELEFAX: 617-720-2441
;
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1128 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
;
; US-08-978-404B-15

```

```

Query Match      8 6%; Score 143.6; DB 2; Length 1128;
Best Local Similarity 52.9%; Pred. No. 6,6e-25;
Matches 438; Conservative 1; Mismatches 360; Indels 29; Gaps 5;

```

```

QY      257 ATTCTGCTTATCTTGATTAATCTCGGTGCGGAGAGAGCGGAGAGAGGCTTC 316
Db      5 AATCTGCTCTGCTGCGCTGCGCTGCGCTGCGGAGAGCGGCTTACGCGGCTCCCA 64
QY      317 TGGGTGTGCGCCCGCCAGACGATCACAGGTGCGAGCATGTCAGTCCGCTCAGTGG 376
Db      65 GGCAGAGCCCTGACAGGATGAGGATCGTGGGGGTCAAGAGGCCCCAGAGCAAGTGG 124
QY      377 CCTGCGAGGTGAGCATACCTATGAAGC-----GTCCATGTGTGTGTGCTGCT 427
Db      125 CCTGCGAGGTGAGCCTGAGATCCAGCGCCATATGAGTCACTTCTGCGGGGCTTC 184
QY      428 CTGCTGTGAGAGAGGTGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 484
Db      185 CTGATCACCCCAAGTGGTGTCTGACCGGAGCATGTGCTGAGACCGGACGTCAAGAT 244
QY      485 AAGAAAGCTATGAGGTCAAGCTGGGGGCCCAACAGTGAATCTTCTTCCAGAGAGCC 544
Db      245 CTGGCGGCTTGAAGGTGCACTGCGGAGAGAGCACTTCTTACACAGAGCAAGCTGTG 304
QY      545 AAGGTAGACACCTGAAGAGACATCACTCCCGCAGCTACTCCAGAGAGGCTCCAG 604
Db      305 CCGGTGAGAGG-----ATCATGTGACACCAAGTTTACACCGCCCAAGATCGA 355
QY      605 GGGCAGATTGACATCTCTCCAACTCAGACAGACCATACCTTCTCCGCTTACATCCGAGCC 664
Db      356 GCGGACATGCGCTGCTGAGAGCTGAGAGAGCCGCTGTAAGTCTCCAGCCACGTCCACAG 415
QY      665 ATCTGCTCTCTGAGCAGCAAGCCTCTTCTCCCAAGCGCTTCACTGCACTGTCTGCTG 724
Db      416 GTCACTCTGCGCTCTCTCTGAGAGCTTCTCCCGGAGATGCTGCTGCTGCTGCTGCTG 475
QY      725 TGGGTGATGAGGCCCCCTCAGTGAAGCTCTGAGCGCCCAAGCACTGAGCAACTGAG 784
Db      476 TGGGGGAGATGAGCAATGATGAGCGCTCCACCGCCATTCTCTGAAAGAGTGAAG 535
QY      785 GTGCTCTGATCACTGCTGAGACGATGTAACCTGCTGTACCAATGACAGCGCAAGCTGA 844
Db      536 GTCCCATATAGAAAAACCACTTTGTGAGCGAAATACCACTTGGCGCTTACAGGGA 595
QY      845 GAGAGCG--CACTTTGTCAAAGAGACATGCTGTGTGCTGCTATGATGAGGGGGCAAG 902
Db      596 GACGACGTCCGATCGCTGCTGAGACATGCTGTGTGCTGCTGCTGCTGCTGCTGCTG 649
QY      903 GAGGCTTGCAGAGGTGATCTGTGGGGGCCCACTTCTGCTGCTGTGAGAGGTCTCTGTAC 962
Db      650 GACTCATGCGAGGACATCTCCGAGGGCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 709
QY      963 CTGACGGGCACTTGTGAGCTGAGGAGATGCTGTGGGGCCGCAACAGGCTGTGTGTAC 1022
Db      710 CAGGCGGGCTGTGAGCTGAGGGGAGAGGCTGTGCGCCAGACCGGCTGCGCATCTAC 769
QY      1023 ACTCTGCTCTCCAGCTATGCTCTCTGATCCAAAGCAAGGTGACAGAA 1070
Db      770 ACCGCTGCTCACTACTGAGCTGATCCACCATATGTCTCCCAA 817

```

```

Search completed: December 15, 2003, 23:34:14
Job time : 114 secs

```





**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

# OM protein - protein search, using sw model

Run on: December 15, 2003, 16:55:33 ; Search time 34 Seconds  
(without alignments)  
994.262 Million cell updates/sec

Title: US-09-925-301-966  
Perfect score: 680  
Sequence: 1 AEVHTKQGEAPAMSGE.....ASNGETLEKTNRPCCVIL 131

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues  
Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0  
Maximum DB seq length: 200000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

## Database :

SPTREMBL\_23:\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_virus:\*  
16: sp\_bacteriap:\*  
17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	597	87.8	115	4 Q9BRT3	Q9BRT3 homo sapien
2	537	79.0	115	11 Q9C086	Q9C086 mus musculu
3	148.5	21.8	95	5 Q9VRA0	Q9VRA0 drosophila
4	137.5	20.2	88	10 Q8BET4	Q8BET4 chlamydomon
5	124.5	18.3	232	10 Q8S227	Q8S227 oryza sativ
6	103.5	15.2	228	10 Q8M1E5	Q8M1E5 arabidopsis
7	99	14.6	92	16 Q8DEL8	Q8DEL8 vibrio vuln
8	93	13.7	93	16 Q9ZS07	Q9ZS07 rhizobium m
9	91.5	13.5	113	11 Q8B159	Q8B159 mus musculu
10	91.5	13.5	1192	11 Q8B179	Q8B179 mus musculu
11	90	13.2	101	16 Q8U1R5	Q8U1R5 archaeobacteri
12	90	13.2	186	10 Q8W465	Q8W465 arabidopsis
13	86	12.6	1332	5 Q9VH2	Q9VH2 drosophila
14	85.5	12.6	367	11 Q91YE3	Q91YE3 mus musculu
15	83	12.2	96	16 Q9HYQ7	Q9HYQ7 pseudomonas
16	83	12.2	179	2 Q33574	Q33574 rhodospirillum rubrum

17	82	12.1	370	16 Q985C6	Q985C6 rhizobium 1
18	82	12.1	469	17 Q9YFC4	Q9YFC4 aeropyrum p
19	81	11.9	97	16 Q9KTC1	Q9KTC1 vibrio chol
20	81	11.9	243	17 Q9YD24	Q9YD24 aeropyrum p
21	81	11.9	541	11 Q8K2W9	Q8K2W9 mus musculu
22	80	11.8	137	4 Q81Y80	Q81Y80 homo sapien
23	80	11.8	137	11 Q8CHV4	Q8CHV4 mus musculu
24	79.5	11.7	605	11 Q8BPK4	Q8BPK4 mus musculu
25	79.5	11.7	1579	11 Q9NMP1	Q9NMP1 mus musculu
26	79.5	11.7	1684	11 Q9WTO5	Q9WTO5 mus musculu
27	79	11.6	395	10 Q23674	Q23674 arabidopsis
28	79	11.6	413	16 Q8KCY4	Q8KCY4 chlorobium
29	78.5	11.5	101	16 Q8YD86	Q8YD86 bruceella me
30	78.5	11.5	101	16 Q8FV19	Q8FV19 bruceella su
31	78.5	11.5	919	11 Q61210	Q61210 mus musculu
32	78.5	11.5	920	11 Q91VL3	Q91VL3 mus musculu
33	78.5	11.5	1856	16 Q97T80	Q97T80 streptococ
34	78	11.5	369	11 Q922Y2	Q922Y2 mus musculu
35	78	11.5	3670	16 Q924X5	Q924X5 streptomyce
36	77.5	11.4	157	16 Q914Q9	Q914Q9 pseudomonas
37	77.5	11.4	844	2 Q9R2J5	Q9R2J5 shigella so
38	77.5	11.4	907	5 Q8M0S4	Q8M0S4 drosophila
39	77.5	11.4	1255	2 Q9R2J6	Q9R2J6 shigella so
40	77	11.3	149	17 Q8TV08	Q8TV08 methanopyru
41	77	11.3	347	16 Q9HXJ3	Q9HXJ3 pseudomonas
42	77	11.3	1028	11 Q07409	Q07409 mus musculu
43	76.5	11.2	166	10 Q49585	Q49585 arabidopsis
44	76.5	11.2	389	17 Q978D6	Q978D6 thermoplasma
45	76.5	11.2	492	17 Q973M8	Q973M8 sulfolobus

## ALIGNMENTS

RESULT 1  
Q9BRT3 PRELIMINARY; PRT; 115 AA.  
ID Q9BRT3  
AC Q9BRT3;  
DT 01-JUN-2001 (TREMREL. 17, Created)  
DT 01-JUN-2001 (TREMREL. 17, Last sequence update)  
DT 01-MAR-2003 (TREMREL. 23, Last annotation update)  
DE Similar to RIKEN cDNA 1810046J19 gene.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Brain;  
RA Strausberg R.;  
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL, BC006006; AAH06006.1; -  
SQ SEQUENCE 115 AA; 12403 MW; 5DBB911C0F23DDC1 CRC64;

Query Match 87.8%; Score 597; DB 4; Length 115;  
Best Local Similarity 100.0%; Pred. No. 2.5e-50;  
Matches 115; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 17 MSGEPCQTSVADPPPEVEPVGSGVRIYVCEPCGEATYLELASAVKEQYPGIIEISRLG 76  
DB 1 MSGEPCQTSVADPPPEVEPVGSGVRIYVCEPCGEATYLELASAVKEQYPGIIEISRLG 60  
QY 77 GTGAFETIENGQLVPSKLENGGPPPEKDIIEAIRASNGETLEKTNRPCCVIL 131  
DB 61 GTGAFETIENGQLVPSKLENGGPPPEKDIIEAIRASNGETLEKTNRPCCVIL 115

RESULT 2  
ID Q9C086 PRELIMINARY; PRT; 115 AA.  
AC Q9C086;  
DT 01-JUN-2001 (TREMREL. 17, Created)  
DT 01-JUN-2001 (TREMREL. 17, Last sequence update)

DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)  
DE 1810046019R1K protein (RIKEN CDNA 1810046019 gene) (Hypothetical  
protein)  
GN 1810046019R1K.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6J; TISSUE=Brain, Pancreas, and Tongue;  
RX MEDLINE=21085660; PubMed=11217851;  
RA Kawai J., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,  
RA Arakawa T., Hara A., Fukunishi Y., Kono H., Adachi J., Fukuda S.,  
RA Azawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana I.,  
RA Saito T., Okazaki Y., Gojibori T., Bono H., Kasukawa T., Saito R.,  
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,  
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,  
RA Kuehl P., Lewis S., Matsuo Y., Nakai I., Pesole G., Quackenbush J.,  
RA Schirml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,  
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barh G.,  
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bona M.F.,  
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,  
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,  
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,  
RA Nordone P., Ring B., Schoenbach C., Rodriguez I., Sakamoto N.,  
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,  
RA Suzuki H., Toyooka K., Wang K.H., Welts C., Whitaker C., Wilming L.,  
RA Wyshak-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohetski S.,  
RA Hayashizaki Y.;  
RT "Functional annotation of a full-length mouse cDNA collection.";  
RL Nature 409:685-690(2001).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Liver;  
RA Strauberg R.;  
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6J; TISSUE=Aorta and vein;  
RX MEDLINE=22354683; PubMed=1246851;  
RA The FANTOM Consortium;  
RA the RIKEN Genome Exploration Research Group Phase I & II Team;  
RT "Analysis of the mouse transcriptome based on functional annotation of  
RT 60,770 full-length cDNAs.";  
RL Nature 420:563-573(2002).  
DR EMBL; AK009922; BAB26586.1; -  
DR EMBL; AK009925; BAB22480.1; -  
DR EMBL; AK007795; BAB25261.1; -  
DR EMBL; BC021589; AAH21589.1; -  
DR EMBL; AK041314; BAC30901.1; -  
DR MGD; MGI:1913678; 1810046019R1K.  
RW Hypothetical protein.  
SQ SEQUENCE 115 AA; 12295 MW; B36A0340DFBA737A CRC64;

Query Match 79.0%; Score 537; DB 11; Length 115;  
Best Local Similarity 89.6%; Pred. No. 1.6e-44;  
Matches 103; Conservative 4; Mismatches 8; Indels 0; Gaps 0;

QY 17 MSGPGQTSVAPPEVEVPGSGVRIIVERYCEPGFEATYTLASAVKQYDGIIESRLG 76  
DB 1 MSGGPAPSVVPPGGEVAGSGVHVYCKPCGFEATYTLASAVKVEYDGIIESRLG 60

QY 77 GTGAFETIENGQVLSKLENGGFPYKDLIAIRRASNGETLEKTNRPCCVLL 131  
DB 61 GTGAFETIENGQVLSKLENGGFPYKDLIAIRRASNGEVEKTNRPCCVLL 115

RESULT 3  
Q9VRAO PRELIMINARY; PRT; 95 AA.  
AC Q9VRAO;  
DT 01-MAY-2000 (TReMBLrel. 13, Created)

DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)  
DE 01-MAR-2003 (TReMBLrel. 23, Last annotation update)  
GN CG15456  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BERKELEY;  
RX MEDLINE=20196006; PubMed=10731132;  
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
RA Brandon R.C., Rogers Y.-H.C., Blazer V.G., Helt G., Nelson C.R., Miklos G.L.G.,  
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
RA Abril J.F., Ashby A., An H.-U., Andrews-Pfankuch C., Baldwin D.,  
RA Ballow R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
RA Borokva D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,  
RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
RA de Pablo S., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,  
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kienison J.A., Ketchum K.A.,  
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
RA Lasok P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
RA Merkulov G., Mlshina N.V., Mobarry C., Morris J., Moshrefi A.,  
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,  
RA Palazolo M., Pittman G.S., Sanders R.D.C., Scheeler F., Shen H.,  
RA Reinert K., Remington K., Sanders R.D.C., Skupski M.P., Smith T.,  
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
RA Splyer R., Spradling A.C., Stapleton M., Strong R., Sun E.,  
RA Stryker R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
RA Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weisenbach J.,  
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
RA Ye J., Yen R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
RA Gibbs R.A., Myers B.W., Rubin G.M., Venter J.C.;  
RT "The genome sequence of Drosophila melanogaster.";  
RL Science 287:2185-2195(2000).  
DR EMBL; AE003571; AAF50903.1; -  
DR FLYBase; FBgn0040650; CG15456.  
SQ SEQUENCE 95 AA; 10506 MW; A4564893BF263FDS CRC64;

Query Match 21.8%; Score 148.5; DB 5; Length 95;  
Best Local Similarity 34.0%; Pred. No. 6.3e-07;  
Matches 32; Conservative 17; Mismatches 44; Indels 1; Gaps 1;

QY 39 VRIIVYCEPCGFEATYTLASAVKQYDGIIESRLGTAETIENGQVLSKLENGG 98  
DB 2 VKVEVEYCGICNPSGQCCHLREFLLASPDLDISCRGRGSEVSDIGQVLSKLSCLA 61

QY 99 FPPEKDLIAIRRASNGETLEKTNRPCCVLL 131  
DB 62 FPGHASTLAQVQAEKGEPEYKDLIAIRRASNGEVEKTNRPCCVLL 95

RESULT 4  
Q8H6T4 PRELIMINARY; PRT; 88 AA.  
AC Q8H6T4;  
DT 01-MAR-2003 (TReMBLrel. 23, Created)

```

DT 01-MAR-2003 (TEMBLrel. 23, last sequence update)
DT 01-MAR-2003 (TEMBLrel. 23, last annotation update)
DE Seloloprotein Selw1.
OS Chlamydomonas reinhardtii.
OC Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
OC Chlamydomonadaceae; Chlamydomonas.
NCBI_TaxID=3055;
RN
RP SEQUENCE FROM N.A.
RA Novoselov S.V., Rao M., Onoshko N.V., Zhi H., Kryukov G.V., Xiang Y.,
RA Weeke D.P., Hatfield D.L., Gladyshev V.N.;
RT "seloloproteins and selencysteine insertion system in the model plant
RT cell system Chlamydomonas reinhardtii."
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF494050; AAN32301.1;
SQ SEQUENCE 88 AA; 9630 MW; DF27CBB4780E1128 CRC64;

Query Match 20.2%; Score 137.5; DB 10; Length 88;
Best Local Similarity 37.2%; Pred. No. 6,7e-06;
Matches 32; Conservative 17; Mismatches 32; Indels 5; Gaps 2;

QY 39 VRIIVEXCEPCGFATYELASVKEQYPGIET---ESRLGAGAFETINQGVSKL 94
Db 4 VQVHVLVCGCGCGTGRSRISLNNIRKFPNADIKFSFEATPQATGFVEVNGELVASKK 63

QY 95 ENGCFYKDLIEAIRRANGETLEK 120
Db 64 NGGGHVDNCKVERI-PAKIGELAK 88

RESULT 5
ID 08S227 PRELIMINARY; PRT; 232 AA.
AC 08S227;
DT 01-JUN-2002 (TEMBLrel. 21, Created)
DT 01-JUN-2002 (TEMBLrel. 21, last sequence update)
DT 01-OCT-2002 (TEMBLrel. 22, last annotation update)
DE P0446G04.25 protein (P0460C04.2 protein).
GN P0446G04.25 OR P0460C04.2.
OS Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Eriarthrideae; Oryzaeae; Oryza.
NCBI_TaxID=39947;
RN
RP SEQUENCE FROM N.A.
RA STRAIN=cv. Nipponbare;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 1, PAC
RT clone: P0446G04."
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
RN
RP SEQUENCE FROM N.A.
RA STRAIN=cv. Nipponbare;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 1, PAC
RT clone: P0446G04."
RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF003252; BAB89601.1;
DR EMBL; AF004366; BAB92910.1;
DR Gramene; Q88227;
SQ SEQUENCE 232 AA; 25385 MW; 7011E960E790E4F CRC64;

Query Match 18.3%; Score 124.5; DB 10; Length 232;
Best Local Similarity 19.9%; Pred. No. 0.00041;
Matches 41; Conservative 21; Mismatches 51; Indels 93; Gaps 4;

QY 9 GPEAPEAAMSG---EPGQTSVAPPEEVE-----PGSGVRIIVEYCE 47
Db 25 GPGQLTPRPDRLPHRPSDAAYVDDIADAAASAOIAEFGVNDGPRASGTTVELKCA 84

QY 48 PCGFATYELASVKEQYPGIET----- 71

```

```

Db 85 SCISYGNNAVTKMKLETSPFGIHLVENVPPPPPKALKAVPELOVGMATLMAGDQIF 144
QY 72 -----ESRLGTGAFETINQGVSKLENGCF 99
Db 145 PRGQWPPWYYSLRANRGCTMATIILFGNPFASFLQSSGAFEVYNGQLVPSKLSQRF 204
QY 100 PYEKDLIEAIR---ASNGETLEK1 121
Db 205 PSEFEELRELIGNRLPDSQFGKNLEKV 230

RESULT 6
ID 08W1E5 PRELIMINARY; PRT; 228 AA.
AC 08W1E5;
DT 01-MAR-2002 (TEMBLrel. 20, Created)
DT 01-MAR-2002 (TEMBLrel. 20, last sequence update)
DT 01-OCT-2002 (TEMBLrel. 22, last annotation update)
DE AT5G58640/man1.90 (Hypothetical protein).
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosid II; Brassicales; Brassicaceae; Arabidopsis.
NCBI_TaxID=3702;
RN
RP SEQUENCE FROM N.A.
RA Shinn P., Chen H., Cheuk R., Kim C.J., Koesema E., Meyers M.C.,
RA Banh J., Bowser L., Carinci P., Chang E., Dale J.M., Goldsmith A.D.,
RA Hayaishizaki Y., Ishida J., Jones T., Kamiya A., Karlin-Neumann G.,
RA Kawai J., Lam B., Lee J.M., Lin J., Miranda M., Narusaka M.,
RA Nguyen M., Onodera C.S., Palm C.J., Quach H.L., Sakurai T., Satou M.,
RA Seki M., Southwick A., Tang C.C., Toriumi M., Wu H.C., Yanada K.,
RA Yamamura Y., Yu G., Yu S., Shinozaki K., Davis R.W., Theologis A.,
RA Ecker J.R.;
RT "Arabidopsis cDNA clones."
RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
RN
RP SEQUENCE FROM N.A.
RA Cheuk R., Chen H., Kim C.J., Meyers M.C., Shinn P., Banh J.,
RA Bowser L., Carinci P., Chang E., Dale J.M., Goldsmith A.D.,
RA Hayaishizaki Y., Ishida J., Jones T., Kamiya A., Karlin-Neumann G.,
RA Kawai J., Lam B., Lee J.M., Lin J., Miranda M., Narusaka M.,
RA Nguyen M., Onodera C.S., Palm C.J., Quach H.L., Sakurai T., Satou M.,
RA Seki M., Southwick A., Tang C.C., Toriumi M., Wu H.C., Yanada K.,
RA Yamamura Y., Yu G., Yu S., Shinozaki K., Davis R.W., Theologis A.,
RA Ecker J.R.;
RT "Arabidopsis ORF clones."
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
RN
RP SEQUENCE FROM N.A.
RA Haas B.J., Volkovskiy N., Town C.D., Troukhan M., Alexandrov N.,
RA Feldmann K.A., Flavell R.B., White O., Salzberg S.L.;
RT "Full-length messenger RNA sequences greatly improve genome
RT annotation."
RL Genome Biol. 0:0-0(2002).
RN
RP SEQUENCE FROM N.A.
RA Brover V., Troukhan M., Alexandrov N., Lu Y.-P., Flavell R.,
RA Feldmann K.;
RT "Full-length cDNA from Arabidopsis thaliana."
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF446358; AAL48231.1;
DR EMBL; AY097419; AAM19935.1;
DR EMBL; AY087582; AAM65124.1;
DR HypoNet; AY087582;
SQ SEQUENCE 228 AA; 24998 MW; E1ACC4135BA6749D CRC64;

Query Match 15.2%; Score 103.5; DB 10; Length 228;
Best Local Similarity 17.9%; Pred. No. 0.044;
Matches 26; Conservative 21; Mismatches 29; Indels 63; Gaps 1;

QY 36 GSGVRIIVEYCEPCGFATYELASVKEQYPGIET----- 71

```

```

DR  EMBL; AL591783; CAC41702.1; -.
KW  Hypothetical protein; Complete proteome.
SQ  SEQUENCE 93 AA; 10528 MW; 8E5590DD60680B2C CRC64;

Query Match
Best Local Similarity 13.7%; Score 93; DB 16; Length 93;
Matches 24; Conservative 17; Mismatches 27; Indels 8; Gaps 4;

QY  40 RIVERCEPCGF--EATYL--ELASAVKEQYPGIEISRLGCTGAFIEINGOLVFSKLE 94
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB  6 RTIILYCTQCNMLLRAGMMAQELLSTFADTLGEVAL--IPGTGAFELRVGALIMWK 62

RESULT 9
Q8BI59 PRELIMINARY; PRT; 1137 AA.
ID 08BI59;
AC 08BI59;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DE 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Weakly similar to KIAA1640 protein.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Oviduct;
RX MEDLINE=22354683; PubMed=12466851;
RA The PANTOM Consortium.
RT "Analysis of the mouse transcriptome based on functional annotation of
RL 60,770 full-length cDNAs."
DR EMBL; AK054172; BAC35680.1; -.
SQ SEQUENCE 1137 AA; 129763 MW; EBD51AF0DF875CBB CRC64;

Query Match
Best Local Similarity 13.5%; Score 91.5; DB 11; Length 1137;
Matches 34; Conservative 16; Mismatches 51; Indels 23; Gaps 4;

QY  15 AAMSGERGQTSVAPPEEVE-----PGSGVRIVEY-----CEPGCFEATYLELASAVK 63
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB  83 AISEGEVGSNGETPPEETVEFIEGTAPDIDVEFISGCTDVEPGEISIQETEVES-IG 141

QY  64 EGYPGIEISRLGCTGAFIEINGOLVFSKLENGGFPYKDLIEAIRRANGSETLEKTN 123
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB  142 EATFGMDVE-----PIKKTMTLNVESIGETSETDVDSIRKLLRGIDLSITV 190

QY  124 SRPP 127
    : : : : :
DB  191 AVPP 194

RESULT 10
Q8BI79 PRELIMINARY; PRT; 1192 AA.
ID 08BI79;
AC 08BI79;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Weakly similar to KIAA1640 protein.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Cerebellum;

```

```

DR  EMBL; AL591783; CAC41702.1; -.
KW  Hypothetical protein; Complete proteome.
SQ  SEQUENCE 93 AA; 10528 MW; 8E5590DD60680B2C CRC64;

Query Match
Best Local Similarity 31.6%; Score 93; DB 16; Length 93;
Matches 24; Conservative 17; Mismatches 27; Indels 8; Gaps 4;

QY  40 RIVERCEPCGF--EATYL--ELASAVKEQYPGIEISRLGCTGAFIEINGOLVFSKLE 94
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB  6 RTIILYCTQCNMLLRAGMMAQELLSTFADTLGEVAL--IPGTGAFELRVGALIMWK 62

RESULT 9
Q8BI59 PRELIMINARY; PRT; 1137 AA.
ID 08BI59;
AC 08BI59;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Weakly similar to KIAA1640 protein.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Oviduct;
RX MEDLINE=22354683; PubMed=12466851;
RA The PANTOM Consortium.
RT "Analysis of the mouse transcriptome based on functional annotation of
RL 60,770 full-length cDNAs."
DR EMBL; AK054172; BAC35680.1; -.
SQ SEQUENCE 1137 AA; 129763 MW; EBD51AF0DF875CBB CRC64;

Query Match
Best Local Similarity 27.4%; Score 91.5; DB 11; Length 1137;
Matches 34; Conservative 16; Mismatches 51; Indels 23; Gaps 4;

QY  15 AAMSGERGQTSVAPPEEVE-----PGSGVRIVEY-----CEPGCFEATYLELASAVK 63
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB  83 AISEGEVGSNGETPPEETVEFIEGTAPDIDVEFISGCTDVEPGEISIQETEVES-IG 141

QY  64 EGYPGIEISRLGCTGAFIEINGOLVFSKLENGGFPYKDLIEAIRRANGSETLEKTN 123
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB  142 EATFGMDVE-----PIKKTMTLNVESIGETSETDVDSIRKLLRGIDLSITV 190

QY  124 SRPP 127
    : : : : :
DB  191 AVPP 194

RESULT 10
Q8BI79 PRELIMINARY; PRT; 1192 AA.
ID 08BI79;
AC 08BI79;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Weakly similar to KIAA1640 protein.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Cerebellum;

```

RX MEDLINE=22354683; PubMed=12466851;  
 RA The FANTOM Consortium;  
 RA The RIKEN Genome Exploration Research Group Phase I & II Team;  
 RT "Analysis of the mouse transcriptome based on functional annotation of  
 RT 60,770 full-length cDNAs."  
 RL Nature 420:563-573(2002).  
 DR EMBL, AK046971; BAC32932.1; -  
 SQ SEQUENCE 1192 AA; 136772 MW; A3BHC5BA0183A3A CRC64;

Query Match 13.5%; Score 91.5; DB 11; Length 1192;  
 Best Local Similarity 27.4%; Pred. No. 5.1;  
 Matches 34; Conservative 16; Mismatches 51; Indels 23; Gaps 4;

QY 15 AAMSGEPGQTSVAPPEEVE-----PGSGVRIVEY-----CEPGFEATYLELA5AVK 63  
 DB 13 AISEGVEGNGETPEPETEVEFGETAPDPTDVEFGETSPGTGDPVETGESIQETVE5-IG 71  
 QY 64 EGYGIGTIESRLGTGAFEIRHNGQLVPSKLENGGFPEKDLIAIRASNGETLEKTN 123  
 DB 72 EATPGMDVE-----PIKKTWELNVE5IGSETSDVD5IRKALRGIDLESITY 120  
 QY 124 SRPP 127  
 DB 121 AYPP 124

## RESULT 11

Q8UR15 PRELIMINARY; PRT; 101 AA.  
 AC Q8UR15;  
 DT 01-JUN-2002 (TrEMBLrel. 21, Created)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
 DE Hypothetical protein Adu0228.  
 GN Adu0228 OR AGR C.387.  
 OS Agrobacterium tumefaciens (strain C58 / ATCC 33970).  
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;  
 OC Rhizobiaceae; Rhizobium.  
 NCBI\_Taxid=176299;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=2160855; PubMed=11743193;  
 RA Wood D.W., Secudal J.C., Kaul R., Monks D.E., Kitajima J.P.,  
 RA Okura V.K., Zhou Y., Chen L., Wood G.E., Almeida N.F., Jr., Woo L.,  
 RA Chen Y., Paulsen I.T., Eilen J.A., Karp P.D., Boyce D. Sr.,  
 RA Chapman P., Clendenen J., Deatherage G., Gillet W., Grant C.,  
 RA Kutayin T., Levy R., Li M.-J., McClelland E., Palmeri A.,  
 RA Raymond C., Rouse G., Saenphimmachak C., Wu Z., Romero P., Gordon D.,  
 RA Zhang S., Yoo H., Tao Y., Biddle P., Jung M., Krespan W., Perry M.,  
 RA Gordon-Kamm B., Liao P., Kim S., Hendrick C., Zhao Z.-Y., Dolan M.,  
 RA Chumley F., Tingey S.V., Tomb J.-F., Gordon M.P., Olson M.V.,  
 RA Neeter B.W.;  
 RT "The genome of the natural genetic engineer Agrobacterium tumefaciens  
 RT C58."  
 RL Science 294:2317-2323(2001).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=2160855; PubMed=11743194;  
 RA Goodner B., Hinkle G., Gattung S., Miller N., Blanchard M.,  
 RA Gurollo B., Goldman B.S., Cao Y., Akenazi M., Halling C., Mullin J.,  
 RA Houmel K., Gordon B., Vaudin M., Iartchak O., Epp A., Liu F.,  
 RA Wollam C., Allinger M., Doughy D., Scott C., Lappas C., Markelz B.,  
 RA Planagan C., Crowell C., Guron J., Lomo C., Sear C., Strub G.,  
 RA Cielo C., Slater S.;  
 RT "Genome sequence of the plant pathogen and biotechnology agent  
 RT Agrobacterium tumefaciens C58."  
 RL Science 294:2323-2328(2001).  
 DR EMBL, AEO08996; AAL41250.1; -  
 DR EMBL, AEO07963; AAK86045.1; -  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 101 AA; 11510 MW; 10A126494F309D21 CRC64;

Query Match 13.2%; Score 90; DB 16; Length 101;

Best Local Similarity 29.6%; Pred. No. 0.32;  
 Matches 24; Conservative 17; Mismatches 34; Indels 6; Gaps 3;

QY 32 EEPGSGVRIVEYCEPGFEATYLELA5AVKQYPG-TEIESRLGTGA-FEIRHNGQL 89  
 DB 3 ETPD---RAIHYCTGQCNWLRAGMWADEIITQTFASDIGEVSILPSTGSLFETVDGTI 58  
 QY 90 VPSKLENGGFPEKDLIEAIR 110  
 DB 59 IWRKRDGGRPGPKELKQRI 79

## RESULT 12

Q8M465 PRELIMINARY; PRT; 186 AA.  
 AC Q8M465;  
 DT 01-MAR-2002 (TrEMBLrel. 20, Created)  
 DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)  
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)  
 DE Hypothetical 20.7 kDa protein.  
 GN ATG31360.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
 NCBI\_Taxid=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Southwick A., Karlin-Neumann G., Nguyen M., Lam B., Miranda M.,  
 RA Palm C.D., Bowser L., Jones T., Banh J., Carninci P., Chen H.,  
 RA Cheuk R., Chung M.K., Hayashizaki Y., Ishida J., Kamiya A., Kawai J.,  
 RA Kim C., Lin J., Liu S.X., Narusaka M., Pham P.K., Sakano H.,  
 RA Sakurai T., Satou M., Seki M., Shinn P., Yamada K., Shinozaki K.,  
 RA Ecker J., Theologis A., Davis R.W.;  
 RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Southwick A., Karlin-Neumann G., Nguyen M., Lam B., Miranda M.,  
 RA Palm C.D., Bowser L., Jones T., Banh J., Carninci P., Chen H.,  
 RA Cheuk R., Chung M.K., Hayashizaki Y., Ishida J., Kamiya A., Kawai J.,  
 RA Kim C., Lin J., Liu S.X., Narusaka M., Pham P.K., Sakano H.,  
 RA Sakurai T., Satou M., Seki M., Shinn P., Yamada K., Shinozaki K.,  
 RA Ecker J., Theologis A., Davis R.W.;  
 DR EMBL, AY062819; AAL32897.1; -  
 DR EMBL, AY081651; AAM10213.1; -  
 KW Hypothetical protein.  
 SQ SEQUENCE 186 AA; 20662 MW; B445DE2FD6785626 CRC64;

Query Match 13.2%; Score 90; DB 10; Length 186;  
 Best Local Similarity 28.6%; Pred. No. 0.7; Length 186;  
 Matches 28; Conservative 16; Mismatches 50; Indels 4; Gaps 3;

QY 2 EVHTRKQGEAPAPAMSGEPGQTSVAPPEEVEPGSGVRIVEYCEPG-GFEATYLELAS 60  
 DB 64 ETGRAKKGEKEVEPEEVAVEEVEKEPEVEDPRTITVEHCKQNAFTRAIQVE 123  
 QY 61 AYKEQYPG--TEIESRLGTGAFEI-EINGQLVPSKLE 95  
 DB 124 ALEGAVGVTVSLNPKPRRGCFEIREBGGQTFISILE 161

RESULT 13

Q9VYH2 PRELIMINARY; PRT; 1322 AA.  
 AC Q9VYH2;  
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE CG18265 protein.  
 GN CG18265.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 NCBI\_TaxID=7227;  
 [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BERKELEY;  
 RX ADAMS M.D., Celniker S.E., Holt R.A., Evans C.A., Gockayne J.D.,  
 RA Amaratilake P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Heit G., Nelson C.R., Miklos G.L.G.,  
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
 RA Baller R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Bernier B.P., Bhandari D., Bolshakov S.,  
 RA Borokova D., Botchan M.R., Bouck H., Brokstein P., Brotler P.,  
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cavley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
 RA Fostler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
 RA Kammel B.E., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Laslo P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Matei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Mekulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Pollard J., Puri V., Reese M.G.,  
 RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yen R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of Drosophila melanogaster";  
 RL Science 287:2185-2195(2000).  
 DR EMBL: AE003524; AAF9339.1;  
 DR FLYBase; FBgn0036725; CG18265.  
 DR InterPro; IPR007087; Znf\_C2H2.  
 DR Pfam; PF00096; zf-C2H2; 9.  
 DR SMART; SM00355; Znf\_C2H2; 10.  
 DR PROSITE; PS00026; ZINC\_FINGER\_C2H2\_1; 8.  
 DR PROSITE; PS50157; ZINC\_FINGER\_C2H2\_2; 3.  
 KW Metal-binding; Zinc; Zinc-finger.  
 SQ SEQUENCE 1322 AA; 143906 MW; 262642D18A5B89 CRC64;  
 Query Match 12.6%; Score 86; DB 5; Length 1322;  
 Best Local Similarity 32.3%; Pred. No. 20;  
 Matches 31; Conservative 8; Mismatches 31; Indels 26; Gaps 5;  
 QY 1 AEVHTKRG-----PEAEPAMSGEPQGSVAP-----PPEVEPGSGVRIV 43  
 DB 820 AEVHGRNGGSGPLRGFTPERPVPASGPG-RTVADAARPPVTLTPTS----- 872  
 QY 44 EYCEPCGFATYLELASAVKQYPIET--SRGG 77  
 DB 873 SYCEICNKELCNKYFMKTHQMGMGIEIENGAIQIG 908  
 RESULT 14  
 ID Q91YE3 PRELIMINARY; PRT; 367 AA.  
 AC Q91YE3;  
 DT 01-DEC-2001 (TREMBlrel. 19, Created)

DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)  
 DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)  
 DE EGIN1 protein (Fragment).  
 GN EGIN1.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 NCBI\_TaxID=10090;  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=21458562; PubMed=11574160;  
 RA Taylor M.S.;  
 RT "Characterisation and comparative analysis of the EGIN gene family";  
 RL Gene 275:125-132(2001).  
 DR EMBL: AJ310546; CAC42515.1; -  
 DR MGD; MGI:1932286; Egin1.  
 DR InterPro; IPR005123; 2OG-Fell\_Oxy.  
 DR InterPro; IPR006620; Pro\_4\_Hyd\_alph.  
 DR InterPro; IPR002893; Znf\_MYND.  
 DR Pfam; PF03171; 2OG-Fell\_Oxy; 1.  
 DR Pfam; PF01753; zf-MYND; 1.  
 DR SMART; SM00702; PAHC; 1.  
 FT NON\_TER  
 SQ SEQUENCE 367 AA; 39551 MW; 87810B3A87A95B55 CRC64;  
 Query Match 12.6%; Score 85.5; DB 11; Length 367;  
 Best Local Similarity 24.4%; Pred. No. 4.5;  
 Matches 40; Conservative 18; Mismatches 45; Indels 61; Gaps 9;  
 QY 5 TRKGPEAPAMSG-----EPG-----QTSVAP-----PPEVEPGSGVR 40  
 DB 68 SRVPGEDDAQAASGSGPAPGSEDPPLSRSPGPERASLCPPAGGGGALSPGGRLPVG 127  
 QY 41 -----IVRYCEPCGFATYLELASAVKQYPIETSLG-----GTGAFEIE 84  
 DB 128 QTPPLPALKALAYIVPC-----MNKGICVDDFLGRTGQIGEVVALHDTKFF--- 179  
 QY 85 INGOLVSKLENGPFYEKDLIEAIRASNGETLEKITNSRPPC 128  
 DB 180 TDQIVSQKSDS-----SKDI-----RGQDITWIEGEPCC 210  
 RESULT 15  
 ID Q9HY07 PRELIMINARY; PRT; 96 AA.  
 AC Q9HY07;  
 DT 01-MAR-2001 (TREMBlrel. 16, Created)  
 DT 01-MAR-2001 (TREMBlrel. 18, Last sequence update)  
 DT 01-OCT-2001 (TREMBlrel. 18, Last annotation update)  
 DE Hypothetical protein PA3338.  
 GN PA3338.  
 OS Pseudomonas aeruginosa.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;  
 OC Pseudomonadaceae; Pseudomonas.  
 NCBI\_TaxID=287;  
 RP SEQUENCE FROM N.A.  
 RX STRAIN=ATCC 15692 / PA01;  
 RX MEDLINE=20437337; PubMed=10984043;  
 RA Stover C.K., Pham X.-O.T., Erwin A.L., Micooguchi S.D., Warren P.,  
 RA Hickey M.J., Brinkman F.S.L., Huftagle W.O., Kowalik D.J., Lagrou M.,  
 RA Garber R.L., Goltzy L., Tolentino E., Westbrook-Wadman S., Yuan Y.,  
 RA Brody L.L., Coulter S.N., Folger K.R., Kae A., Larbig K., Lim R.M.,  
 RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,  
 RA Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;  
 RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pathogen.";  
 RL Nature 406:959-964(2000).  
 DR EMBL: AE004756; AAG06726.1; -  
 KW Hypothetical protein; Complete Proteome.  
 SQ SEQUENCE 96 AA; 10812 MW; 3ACC280DC71A1BC3 CRC64;  
 Query Match 12.2%; Score 83; DB 16; Length 96;



Best Local Similarity 31.2%; Pred. No. 1.5;  
Matches 25; Conservative 13; Mismatches 36; Indels 6; Gaps 3;

Matches 25; Conservative 13; Mismatches 36; Indels 6; Gaps 3;

**QY** 35 PGSGRIVERCEPCGF--EAYLL-ELASVAHQYPGIEISRIIGTGAFFIEINQLV 90  
| : | | : | : | : | : | : | : | : | : | : | : | : | : | : |  
**Ddb** 2 PFAKEIVITCTQCQWLLRAAWLAQELLSTFADDLGKVCLEPFG--GVFRITCDGVQV 59

Db 2 PTAKEPIVITYCTQCQMLRAWLAQELLSTFADDLGKVCLEPGTG--GVFRITCDGNQV 59

QY 91 FSKLENGGFYKDLIEAIR 110

```

Db      : : |||| | | : :
        60 WERKADGGFPEAKALKQRRV 79

```

Search completed: December 15, 2003, 16:59:05  
Job time : 39 secs

Job time : 39 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 15, 2003, 16:59:11 ; Search time 31 Seconds

(without alignments)  
785.931 Million cell updates/sec

Title: US-09-925-301-966

Perfect score: 680  
Sequence: 1 AEVHTKQGEAPAMSGE.....ASNGETLEKITSRPPCVIL 131

Scoring table: BLOSUM62  
Gap 10.0 , Gapext 0.5

Searched: 684280 seqs, 185983659 residues

Total number of hits satisfying chosen parameters: 684280

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :

Published Applications\_AA:\*

- 1: /cgn2\_6/ptodata/1/pubppaa/US07\_PUBCOMB.pep.\*
- 2: /cgn2\_6/ptodata/1/pubppaa/PCR\_NEW\_PUB.pep.\*
- 3: /cgn2\_6/ptodata/1/pubppaa/US06\_NEW\_PUB.pep.\*
- 4: /cgn2\_6/ptodata/1/pubppaa/US06\_PUBCOMB.pep.\*
- 5: /cgn2\_6/ptodata/1/pubppaa/US07\_NEW\_PUB.pep.\*
- 6: /cgn2\_6/ptodata/1/pubppaa/PCR\_NEW\_PUBCOMB.pep.\*
- 7: /cgn2\_6/ptodata/1/pubppaa/US08\_NEW\_PUB.pep.\*
- 8: /cgn2\_6/ptodata/1/pubppaa/US08\_PUBCOMB.pep.\*
- 9: /cgn2\_6/ptodata/1/pubppaa/US09\_PUBCOMB.pep.\*
- 10: /cgn2\_6/ptodata/1/pubppaa/US09\_PUBCOMB.pep.\*
- 11: /cgn2\_6/ptodata/1/pubppaa/US09C\_PUBCOMB.pep.\*
- 12: /cgn2\_6/ptodata/1/pubppaa/US09\_NEW\_PUB.pep.\*
- 13: /cgn2\_6/ptodata/1/pubppaa/US10\_PUBCOMB.pep.\*
- 14: /cgn2\_6/ptodata/1/pubppaa/US10C\_PUBCOMB.pep.\*
- 15: /cgn2\_6/ptodata/1/pubppaa/US10C\_PUBCOMB.pep.\*
- 16: /cgn2\_6/ptodata/1/pubppaa/US10\_NEW\_PUB.pep.\*
- 17: /cgn2\_6/ptodata/1/pubppaa/US60\_NEW\_PUB.pep.\*
- 18: /cgn2\_6/ptodata/1/pubppaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	680	100.0	131	9	US-09-925-301-966
2	626	92.1	206	15	US-10-177-293-480
3	605	89.0	117	12	US-09-833-203-34
4	597	87.8	115	10	US-09-824-787B-2
5	80	11.8	137	11	US-09-374-046A-96
6	80	11.8	146	11	US-09-892-877-301
7	80	11.8	146	11	US-09-948-783-314
8	77	11.3	1028	11	US-09-947-063-14
9	76.5	11.2	298	15	US-10-156-761-9483
10	76	11.2	425	10	US-09-764-877-1163
11	76	11.2	425	10	US-09-860-670-105
12	74	10.9	772	10	US-09-909-320-339
13	74	10.9	772	10	US-09-908-088B-339
14	74	10.9	772	10	US-09-905-291A-339
15	74	10.9	772	10	US-09-902-853-339

16	74	10.9	772	10	US-09-907-824-339	Sequence 339, App
17	74	10.9	772	10	US-09-907-841-339	Sequence 339, App
18	74	10.9	772	10	US-09-904-011-339	Sequence 339, App
19	74	10.9	772	11	US-09-906-742-339	Sequence 339, App
20	74	10.9	772	11	US-09-905-838-339	Sequence 339, App
21	74	10.9	772	11	US-09-907-613-339	Sequence 339, App
22	74	10.9	772	11	US-09-907-942-339	Sequence 339, App
23	74	10.9	772	11	US-09-904-859-339	Sequence 339, App
24	74	10.9	772	11	US-09-909-204-339	Sequence 339, App
25	74	10.9	772	11	US-09-904-820-339	Sequence 339, App
26	74	10.9	772	11	US-09-904-786-339	Sequence 339, App
27	74	10.9	772	11	US-09-906-646-339	Sequence 339, App
28	74	10.9	772	11	US-09-906-700-339	Sequence 339, App
29	74	10.9	772	11	US-09-903-786-339	Sequence 339, App
30	74	10.9	772	11	US-09-902-903-339	Sequence 339, App
31	74	10.9	772	11	US-09-903-749A-339	Sequence 339, App
32	74	10.9	772	11	US-09-904-119-339	Sequence 339, App
33	74	10.9	772	11	US-09-904-986-339	Sequence 339, App
34	74	10.9	772	11	US-09-902-736-339	Sequence 339, App
35	74	10.9	772	11	US-09-907-794-339	Sequence 339, App
36	74	10.9	772	11	US-09-903-943-339	Sequence 339, App
37	74	10.9	772	11	US-09-904-462-339	Sequence 339, App
38	74	10.9	772	11	US-09-907-925-339	Sequence 339, App
39	74	10.9	772	11	US-09-902-692-339	Sequence 339, App
40	74	10.9	772	11	US-09-903-520-339	Sequence 339, App
41	74	10.9	772	11	US-09-905-056-339	Sequence 339, App
42	74	10.9	772	11	US-09-909-064-339	Sequence 339, App
43	74	10.9	772	11	US-09-904-553-339	Sequence 339, App
44	74	10.9	772	11	US-09-905-381-339	Sequence 339, App
45	74	10.9	772	11	US-09-905-088-339	Sequence 339, App

#### ALIGNMENTS

RESULT 1  
US-09-925-301-966  
; Sequence 966, Application US/09925301  
; Patent No. US20020052308A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies  
; FILE REFERENCE: PA106  
; CURRENT APPLICATION NUMBER: US/09/925,301  
; CURRENT FILING DATE: 2001-08-10  
; PRIOR APPLICATION NUMBER: PCT/US00/05882  
; PRIOR FILING DATE: 2000-03-08  
; PRIOR APPLICATION NUMBER: 60/124,270  
; PRIOR FILING DATE: 1999-03-12  
; NUMBER OF SEQ ID NOS: 1694  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 966  
; LENGTH: 131  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-925-301-966

Query Match 100.0%; Score 680; DB 9; Length 131;  
Best Local Similarity 100.0%; Pred. No. 1.2e-64;  
Matches 131; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	AEVHTKQGEAPAMSGE	PGQTSVAPPPEE	PGSGVRIWVEY	CGGFATYIELAS	60
DB	1	AEVHTKQGEAPAMSGE	PGQTSVAPPPEE	PGSGVRIWVEY	CGGFATYIELAS	60
QY	61	AYVEQYPGIEISRLG	TGTAFFIEIN	QGVFSKLENG	CPYEXDLTEAIRAS	120
DB	61	AYVEQYPGIEISRLG	TGTAFFIEIN	QGVFSKLENG	CPYEXDLTEAIRAS	120
QY	121	ITNSRPPCVIL	131			
DB	121	ITNSRPPCVIL	131			



APPLICANT: Lavallie, Edward R.  
APPLICANT: Collins-Racie, Lisa A.  
APPLICANT: Evans, Cheryl  
APPLICANT: Merberg, David  
APPLICANT: Treacy, Maurice  
APPLICANT: Agostino, Michael J.  
APPLICANT: Steininger II, Robert J.  
APPLICANT: Spaulding, Vilki  
APPLICANT: Wong, Gordon G.  
APPLICANT: Clark, Hilary  
APPLICANT: Fechtel, Kim  
APPLICANT: Genetics Institute, Inc.  
TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES ENCODING THEM  
FILE REFERENCE: GI 6075-83A  
CURRENT APPLICATION NUMBER: US/09/374,046A  
CURRENT FILING DATE: 1999-08-13  
NUMBER OF SEQ ID NOS: 240  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 96  
LENGTH: 137  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-374-046A-96

Query Match 11.8%; Score 80; DB 11; Length 137;  
Best Local Similarity 43.6%; Pred. No. 1.1;  
Matches 17; Conservative 9; Mismatches 13; Indels 0; Gaps 0;

QY 71 IESRLGTAFFIEINGQLVFSKLENGCPPEYEDLIEAI 109  
DB 81 IENQCMSTGAFFITLNDVPVWSKLESGLHPSMOQLVOIL 119

RESULT 6  
US-09-892-877-301  
Sequence 301, Application US/09892877  
Publication No. US20030077809A1  
GENERAL INFORMATION:  
APPLICANT: Ruben et. al.  
TITLE OF INVENTION: 97 Human secreted proteins  
FILE REFERENCE: P2028P1  
CURRENT APPLICATION NUMBER: US/09/892,877  
CURRENT FILING DATE: 2001-06-28  
PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: US/09/437,658  
PRIOR FILING DATE: EARLIER FILING DATE: 1999-11-10  
NUMBER OF SEQ ID NOS: 461  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 301  
LENGTH: 146  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-892-877-301

Query Match 11.8%; Score 80; DB 11; Length 146;  
Best Local Similarity 43.6%; Pred. No. 1.2;  
Matches 17; Conservative 9; Mismatches 13; Indels 0; Gaps 0;

QY 71 IESRLGTAFFIEINGQLVFSKLENGCPPEYEDLIEAI 109  
DB 90 IENQCMSTGAFFITLNDVPVWSKLESGLHPSMOQLVOIL 128

RESULT 7  
US-09-948-783-314  
Sequence 314, Application US/09948783  
Publication No. US20030100051A1  
GENERAL INFORMATION:  
APPLICANT: Ruben et. al.  
TITLE OF INVENTION: 97 Human secreted proteins  
FILE REFERENCE: P2028P2  
CURRENT APPLICATION NUMBER: US/09/948,783  
CURRENT FILING DATE: 2001-09-10  
PRIOR APPLICATION NUMBER: 60/231,846

PRIOR FILING DATE: 2000-09-11  
PRIOR APPLICATION NUMBER: 09/892,877  
PRIOR FILING DATE: 2001-06-28  
PRIOR APPLICATION NUMBER: 09/437,658  
PRIOR FILING DATE: 1998-11-10  
PRIOR APPLICATION NUMBER: PCT/US99/09847  
PRIOR FILING DATE: 1999-05-06  
PRIOR APPLICATION NUMBER: 60/085,093  
PRIOR FILING DATE: 1998-05-12  
PRIOR APPLICATION NUMBER: 60/085,094  
PRIOR FILING DATE: 1998-05-12  
PRIOR APPLICATION NUMBER: 60/085,105  
PRIOR FILING DATE: 1998-05-12  
PRIOR APPLICATION NUMBER: 60/085,180  
PRIOR FILING DATE: 1998-05-12  
PRIOR APPLICATION NUMBER: 60/085,927  
PRIOR FILING DATE: 1998-05-18  
PRIOR APPLICATION NUMBER: 60/085,906  
PRIOR FILING DATE: 1998-05-18  
PRIOR APPLICATION NUMBER: 60/085,924  
PRIOR FILING DATE: 1998-05-18  
PRIOR APPLICATION NUMBER: 60/085,922  
PRIOR FILING DATE: 1998-05-18  
PRIOR APPLICATION NUMBER: 60/085,921  
PRIOR FILING DATE: 1998-05-18  
PRIOR APPLICATION NUMBER: 60/085,923  
PRIOR FILING DATE: 1998-05-18  
PRIOR APPLICATION NUMBER: 60/085,925  
PRIOR FILING DATE: 1998-05-18  
PRIOR APPLICATION NUMBER: 60/085,928  
PRIOR FILING DATE: 1998-05-18  
PRIOR APPLICATION NUMBER: 60/085,920  
PRIOR FILING DATE: 1998-05-18  
NUMBER OF SEQ ID NOS: 465  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 314  
LENGTH: 146  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-948-783-314

Query Match 11.8%; Score 80; DB 11; Length 146;  
Best Local Similarity 43.6%; Pred. No. 1.2;  
Matches 17; Conservative 9; Mismatches 13; Indels 0; Gaps 0;

QY 71 IESRLGTAFFIEINGQLVFSKLENGCPPEYEDLIEAI 109  
DB 90 IENQCMSTGAFFITLNDVPVWSKLESGLHPSMOQLVOIL 128

RESULT 8  
US-09-947-063-14  
Sequence 14, Application US/09947063  
Publication No. US20030059775A1  
GENERAL INFORMATION:  
APPLICANT: Padigan et al.  
TITLE OF INVENTION: No. US20030059775A1 Proteins and Nucleic Acids Encoding Same  
FILE REFERENCE: 21402-112  
CURRENT APPLICATION NUMBER: US/09/947,063  
CURRENT FILING DATE: 2001-09-05  
PRIOR APPLICATION NUMBER: 60/229,990  
PRIOR FILING DATE: 2000-09-05  
PRIOR APPLICATION NUMBER: 60/229,988  
PRIOR FILING DATE: 2000-09-05  
NUMBER OF SEQ ID NOS: 30  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 14  
LENGTH: 1028  
TYPE: PRT  
ORGANISM: Mus musculus  
US-09-947-063-14

Query Match 11.3%; Score 77; DB 11; Length 1028;

Best Local Similarity 24.4%; Pred. No. 34;  
Matches 42; Conservative 21; Mismatches 49; Indels 60; Gaps 10;

QY 2 EVHTRKQGPAPAAAGSGEGQTSVA-----PPPEVEPSSGVRIVVEYCEPGFEATYTL 56

DB 692 KATTEEAAPPIAAEVEGGGSSSELVTWDPVPEELQNGGQGVAVF-RPLGV-TTWI 749

QY 57 E-LASAVKEQY-----PGIEIESRLG-----GTGAF----- 81

DB 750 QTVVTPDNRRYVFRNESIYFPSPYEVKVGNNKGGPSPVTYVSAEEPTVAPSHI 809

QY 82 -----EIEINGQLVFSKLENGG-FPYEKDLIARASNG---ETLEKI 121

DB 810 SAHSLSSEIEVSWNTIPMTLSNGHLGVE-----VRWNNGGEESSSRKV 855

RESULT 9  
US-10-156-761-9483  
Sequence 9483, Application US/10156761

Publication No. US20030119018A1  
GENERAL INFORMATION:

APPLICANT: OMTURA, SATOSHI  
APPLICANT: IKEDA, HARUO

APPLICANT: ISHIKAWA, JUN  
APPLICANT: HORIKAWA, HIROSHI

APPLICANT: SHIBA, TADAYOSHI  
APPLICANT: SAKAKI, YOSHIYUKI

APPLICANT: HATTORI, MASAHIRA  
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES

FILE REFERENCE: 249-262  
CURRENT APPLICATION NUMBER: US/10/156, 761

CURRENT FILING DATE: 2002-05-29  
PRIOR APPLICATION NUMBER: JP 2001-204089

PRIOR FILING DATE: 2001-05-30  
PRIOR APPLICATION NUMBER: JP 2001-272697

PRIOR FILING DATE: 2001-08-02  
NUMBER OF SEQ ID NOS: 15109

SEQ ID NO 9483  
LENGTH: 298

TYPE: PRT  
ORGANISM: Streptomyces avermitilis

US-10-156-761-9483  
Query Match 11.2%; Score 76.5; DB 15; Length 298;

Best Local Similarity 25.5%; Pred. No. 7.3;  
Matches 26; Conservative 14; Mismatches 49; Indels 13; Gaps 4;

QY 6 RKQGPAAE--PAAMSGEGQTSVAPPEEVEPGSGVRIVVEY---CEPCG--FEATYLEL 58

DB 118 RERNPERRPTPSPTRAAPSPTRAPPPASATPGSPVPVQOERGCPSGIRFDSGPDVA 177

QY 59 ASAVKEQYPGIEIESRLGCGTGAPEI-EINGQLVFSKLENGGP 100

DB 178 AMGLR-----AMTLTLTNGCKRSYEVNGYPSVAVLDERGP 213

RESULT 10  
US-09-764-877-1163  
Sequence 1163, Application US/09764877

Patent No. US20020147140A1  
GENERAL INFORMATION:

APPLICANT: Rosen et al.  
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies

FILE REFERENCE: PC005  
CURRENT APPLICATION NUMBER: US/09/764, 877

CURRENT FILING DATE: 2001-01-17  
Prior application data removed - refer to PALM or file wrapper

NUMBER OF SEQ ID NOS: 4031  
SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 1163  
LENGTH: 425  
TYPE: PRT  
ORGANISM: Homo sapiens

FEATURE:  
NAME/KEY: SITE  
LOCATION: (140)

OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids

NAME/KEY: SITE  
LOCATION: (196)

OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids

NAME/KEY: SITE  
LOCATION: (351)

OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids

NAME/KEY: SITE  
LOCATION: (368)

OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids

US-09-764-877-1163  
Query Match 11.2%; Score 76; DB 10; Length 425;

Best Local Similarity 25.2%; Pred. No. 13;  
Matches 26; Conservative 14; Mismatches 55; Indels 8; Gaps 3;

QY 3 VHTRKQGPAPAAAGSGEGQTSVAPPEEVEPGSGVRIVVEY---CEPCGFEATYTL 56

DB 274 VHQGFNPALSPQSRPPGPGAGPDPSPPGADPSRGAFIGRFDQASAGCGFYVADYL 333

QY 57 ELASAVKEQYPGIEIESRLGCGTGAPEI-EINGQLVFSKLENG 97

DB 334 AARALAGELAGQEEEXLEGLEVMDFLRFGSLXLFRAVEPG 376

RESULT 11  
US-09-860-670-105  
Sequence 105, Application US/09860670

Patent No. US20020165137A1  
GENERAL INFORMATION:

APPLICANT: Ruben et al.  
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies

FILE REFERENCE: PA127P1  
CURRENT APPLICATION NUMBER: US/09/860, 670

CURRENT FILING DATE: 2001-05-21  
Prior application data removed - consult PALM or file wrapper

NUMBER OF SEQ ID NOS: 289  
SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 105  
LENGTH: 425

TYPE: PRT  
ORGANISM: Homo sapiens

FEATURE:  
NAME/KEY: SITE  
LOCATION: (140)

OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids

NAME/KEY: SITE  
LOCATION: (196)

OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids

NAME/KEY: SITE  
LOCATION: (351)

OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids

NAME/KEY: SITE  
LOCATION: (368)

OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids

US-09-860-670-105  
Query Match 11.2%; Score 76; DB 10; Length 425;

Best Local Similarity 25.2%; Pred. No. 13;  
Matches 26; Conservative 14; Mismatches 55; Indels 8; Gaps 3;

QY 3 VHTRKQGPAPAAAGSGEGQTSVAPPEEVEPGSGVRIVVEY---CEPCGFEATYTL 56

DB 274 VHQGFNPALSPQSRPPGPGAGPDPSPPGADPSRGAFIGRFDQASAGCGFYVADYL 333

QY 57 ELASAVKEQYPGIEIESRLGCGTGAPEI-EINGQLVFSKLENG 97

DB 334 AARALAGELAGQEEEXLEGLEVMDFLRFGSLXLFRAVEPG 376

RESULT 12  
US-09-909-320-339  
Sequence 339, Application US/09909320  
Patent No. US20020132240A1  
GENERAL INFORMATION:  
APPLICANT: Genentech, Inc.  
APPLICANT: Ashkenazi, Avi  
APPLICANT: Botstein, David  
APPLICANT: Desnoyers, Luc  
APPLICANT: Eaton, Dan L.  
APPLICANT: Ferrara, Napoleone  
APPLICANT: Filvaroff, Ellen  
APPLICANT: Fong, Sherman  
APPLICANT: Gao, Wei-Qiang  
APPLICANT: Gerber, Hanspeter  
APPLICANT: Gertschen, Mary E.  
APPLICANT: Goddard, A.  
APPLICANT: Godowski, Paul J.  
APPLICANT: Grimaldi, Christopher J.  
APPLICANT: Gurney, Austin L.  
APPLICANT: Hillan, Kenneth, J.  
APPLICANT: Kljavin, Ivar J.  
APPLICANT: Mather, Jennie P.  
APPLICANT: Pan, James  
APPLICANT: Paoni, Nicholas F.  
APPLICANT: Roy, Margaret Ann  
APPLICANT: Stewart, Timothy A.  
APPLICANT: Tumas, Daniel  
APPLICANT: Williams, P. Mickey  
APPLICANT: Wood, William, I.  
TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic  
FILE REFERENCE: 10466-14  
CURRENT APPLICATION NUMBER: US/09/909,320  
PRIOR FILING DATE: 2002-01-04  
PRIOR APPLICATION NUMBER: PCT/US00/04414  
PRIOR FILING DATE: 2000-02-22  
PRIOR APPLICATION NUMBER: US 60/143,048  
PRIOR FILING DATE: 1999-07-07  
PRIOR APPLICATION NUMBER: US 60/145,698  
PRIOR FILING DATE: 1999-07-26  
PRIOR APPLICATION NUMBER: US 60/146,222  
PRIOR FILING DATE: 1999-07-28  
PRIOR APPLICATION NUMBER: PCT/US99/20594  
PRIOR FILING DATE: 1999-09-08  
PRIOR APPLICATION NUMBER: PCT/US99/20944  
PRIOR FILING DATE: 1999-09-13  
PRIOR APPLICATION NUMBER: PCT/US99/21090  
PRIOR FILING DATE: 1999-09-15  
PRIOR APPLICATION NUMBER: PCT/US99/21547  
PRIOR FILING DATE: 1999-09-15  
PRIOR APPLICATION NUMBER: PCT/US99/23089  
PRIOR FILING DATE: 1999-10-05  
PRIOR APPLICATION NUMBER: PCT/US99/28214  
PRIOR FILING DATE: 1999-11-29  
PRIOR APPLICATION NUMBER: PCT/US99/28313  
PRIOR FILING DATE: 1999-11-30  
PRIOR APPLICATION NUMBER: PCT/US99/28564  
PRIOR FILING DATE: 1999-12-02  
PRIOR APPLICATION NUMBER: PCT/US99/28565  
PRIOR FILING DATE: 1999-12-02  
PRIOR APPLICATION NUMBER: PCT/US99/30095  
PRIOR FILING DATE: 1999-12-16  
PRIOR APPLICATION NUMBER: PCT/US99/30911  
PRIOR FILING DATE: 1999-12-20  
PRIOR APPLICATION NUMBER: PCT/US99/30999  
PRIOR FILING DATE: 1999-12-20  
PRIOR APPLICATION NUMBER: PCT/US00/00219  
PRIOR FILING DATE: 2000-01-05  
NUMBER OF SEQ ID NOS: 423  
SEQ ID NO 339  
LENGTH: 772  
TYPE: PRT

ORGANISM: Homo Sapien  
US-09-909-320-339  
Query Match 10.9%; Score 74; DE 10; Length 772;  
Best Local Similarity 25.2%; Pred. No. 49;  
Matches 26; Conservative 14; Mismatches 55; Indels 8; Gaps 3;  
Qy 3 VTRKQPEAP--AASGEPQTSVAPPEVEPGSGVRIVY-----CEPCFATYL 56  
Db 621 VHFQENPLSPRSPPPGAGDPDPSPPGADPSRGAPICGRFDQASAGCFYNADYL 680  
Qy 57 ELASAVEQYPGIEISRLGTCAPFI--EINGQLVPSKLENG 97  
Db 681 AARARLAGEIAGQEEEALEGLEVMDFLRFSGLHFRFAVEPG 723  
RESULT 13  
US-09-909-088B-339  
Sequence 339, Application US/09909088B  
Patent No. US20020146709A1  
GENERAL INFORMATION:  
APPLICANT: Genentech, Inc.  
APPLICANT: Ashkenazi, Avi  
APPLICANT: Botstein, David  
APPLICANT: Desnoyers, Luc  
APPLICANT: Eaton, Dan L.  
APPLICANT: Ferrara, Napoleone  
APPLICANT: Filvaroff, Ellen  
APPLICANT: Fong, Sherman  
APPLICANT: Gao, Wei-Qiang  
APPLICANT: Gerber, Hanspeter  
APPLICANT: Gertschen, Mary E.  
APPLICANT: Goddard, A.  
APPLICANT: Godowski, Paul J.  
APPLICANT: Grimaldi, Christopher J.  
APPLICANT: Gurney, Austin L.  
APPLICANT: Hillan, Kenneth, J.  
APPLICANT: Kljavin, Ivar J.  
APPLICANT: Mather, Jennie P.  
APPLICANT: Pan, James  
APPLICANT: Paoni, Nicholas F.  
APPLICANT: Roy, Margaret Ann  
APPLICANT: Stewart, Timothy A.  
APPLICANT: Tumas, Daniel  
APPLICANT: Williams, P. Mickey  
APPLICANT: Wood, William, I.  
TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic  
FILE REFERENCE: 10466-14  
CURRENT APPLICATION NUMBER: US/09/909,088B  
PRIOR FILING DATE: 2001-07-18  
PRIOR APPLICATION NUMBER: PCT/US00/04414  
PRIOR FILING DATE: 2000-02-22  
PRIOR APPLICATION NUMBER: US 60/143,048  
PRIOR FILING DATE: 1999-07-07  
PRIOR APPLICATION NUMBER: US 60/145,698  
PRIOR FILING DATE: 1999-07-26  
PRIOR APPLICATION NUMBER: US 60/146,222  
PRIOR FILING DATE: 1999-07-28  
PRIOR APPLICATION NUMBER: PCT/US99/20594  
PRIOR FILING DATE: 1999-09-08  
PRIOR APPLICATION NUMBER: PCT/US99/20944  
PRIOR FILING DATE: 1999-09-13  
PRIOR APPLICATION NUMBER: PCT/US99/21090  
PRIOR FILING DATE: 1999-09-15  
PRIOR APPLICATION NUMBER: PCT/US99/21547  
PRIOR FILING DATE: 1999-09-15  
PRIOR APPLICATION NUMBER: PCT/US99/23089  
PRIOR FILING DATE: 1999-10-05  
PRIOR APPLICATION NUMBER: PCT/US99/28214  
PRIOR FILING DATE: 1999-11-29  
PRIOR APPLICATION NUMBER: PCT/US99/28313  
PRIOR FILING DATE: 1999-11-30



```

; PRIOR APPLICATION NUMBER: PCT/US99/28564
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/28565
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/30095
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: PCT/US99/30911
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US99/30999
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US00/00219
; PRIOR FILING DATE: 2000-01-05
; NUMBER OF SEQ ID NOS: 423
; SEQ ID NO 339
; LENGTH: 772
; TYPE: PRT
; ORGANISM: Homo Sapien
; US-09-909-088B-339
```

```

Query Match      10.9%; Score 74; DB 10; Length 772;
Best Local Similarity 25.2%; Pred. No. 49;
Matches 26; Conservative 14; Mismatches 55; Indels 8; Gaps 3;
```

```

QY      3 VHTRKQGEAP-AAMSGEPQTSVAPPEVEPGSGVRIVVEY-----CEPCGFATYLL 56
        |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db      621 VHFQEFNPALSPQSPGPPGAGDPSPGADPSRGATIGRFDQASAEQCFYNADYL 680
QY      57 ELASAVEQYPGIIEISRLGTTGAFEL--EINGQVFSKLENG 97
        : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      681 AARARLAGELAGQEEBALGLEVMDFLRFSGHLFRFAVEPG 723
```

## RESULT 14

```

; Sequence 339, Application US/09905291A
; Patent No. US20020160374A1
; GENERAL INFORMATION:
; APPLICANT: Genentech, Inc.
; APPLICANT: Aeshkenazi, Avi
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan L.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Fong, Sherman
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerber, Hanspeter
; APPLICANT: Gottfredsen, Mary E.
; APPLICANT: Grimaldi, Paul J.
; APPLICANT: Gueney, Austin L.
; APPLICANT: Hillan, Kenneth, J.
; APPLICANT: Kijavlin, Ivar J.
; APPLICANT: Mather, Jennie P.
; APPLICANT: Pan, James
; APPLICANT: Paoni, Nicholas F.
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Williams, P. Mickey
; APPLICANT: Wood, William, I.
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; TITLE OF INVENTION: Acids Encoding the Same
; FILE REFERENCE: 10466-14
; CURRENT APPLICATION NUMBER: US/09/905, 291A
; PRIOR APPLICATION NUMBER: PCT/US00/04414
; PRIOR FILING DATE: 2000-02-22
; PRIOR APPLICATION NUMBER: US 60/143, 048
; PRIOR FILING DATE: 1999-07-07
; PRIOR APPLICATION NUMBER: US 60/145, 698
; PRIOR FILING DATE: 1999-07-26
```

```

; PRIOR APPLICATION NUMBER: US 60/146, 222
; PRIOR FILING DATE: 1999-07-28
; PRIOR APPLICATION NUMBER: PCT/US99/20594
; PRIOR FILING DATE: 1999-09-08
; PRIOR APPLICATION NUMBER: PCT/US99/20944
; PRIOR FILING DATE: 1999-09-13
; PRIOR APPLICATION NUMBER: PCT/US99/21090
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/21547
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/23089
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: PCT/US99/28214
; PRIOR FILING DATE: 1999-11-29
; PRIOR APPLICATION NUMBER: PCT/US99/28313
; PRIOR FILING DATE: 1999-11-30
; PRIOR APPLICATION NUMBER: PCT/US99/28564
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/28565
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/30095
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: PCT/US99/30911
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US99/30999
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US00/00219
; PRIOR FILING DATE: 2000-01-05
; NUMBER OF SEQ ID NOS: 423
; SEQ ID NO 339
; LENGTH: 772
; TYPE: PRT
; ORGANISM: Homo Sapien
; US-09-905-291A-339
```

```

Query Match      10.9%; Score 74; DB 10; Length 772;
Best Local Similarity 25.2%; Pred. No. 49;
Matches 26; Conservative 14; Mismatches 55; Indels 8; Gaps 3;
```

```

QY      3 VHTRKQGEAP-AAMSGEPQTSVAPPEVEPGSGVRIVVEY-----CEPCGFATYLL 56
        |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db      621 VHFQEFNPALSPQSPGPPGAGDPSPGADPSRGATIGRFDQASAEQCFYNADYL 680
QY      57 ELASAVEQYPGIIEISRLGTTGAFEL--EINGQVFSKLENG 97
        : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      681 AARARLAGELAGQEEBALGLEVMDFLRFSGHLFRFAVEPG 723
```

## RESULT 15

```

; Sequence 339, Application US/09902853
; Publication No. US20020192659A1
; GENERAL INFORMATION:
; APPLICANT: Genentech, Inc.
; APPLICANT: Aeshkenazi, Avi
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan L.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Fong, Sherman
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerber, Hanspeter
; APPLICANT: Gottfredsen, Mary E.
; APPLICANT: Grimaldi, Paul J.
; APPLICANT: Gueney, Austin L.
; APPLICANT: Hillan, Kenneth, J.
; APPLICANT: Kijavlin, Ivar J.
; APPLICANT: Mather, Jennie P.
; APPLICANT: Pan, James
; APPLICANT: Paoni, Nicholas F.
```

```

/ APPLICANT: Roy, Margaret Ann
/ APPLICANT: Stewart, Timothy A.
/ APPLICANT: Tumas, Daniel
/ APPLICANT: Williams, P. Mickey
/ APPLICANT: Wood, William, I.
/ TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
/ TITLE OF INVENTION: Acids Encoding the Same
/ FILE REFERENCE: 10466-14
/ CURRENT APPLICATION NUMBER: US/09/902,853
/ CURRENT FILING DATE: 2001-07-10
/ PRIOR APPLICATION NUMBER: US/09/665,350
/ PRIOR FILING DATE: 2000-09-18
/ PRIOR APPLICATION NUMBER: US 60/143,048
/ PRIOR FILING DATE: 1999-07-07
/ PRIOR APPLICATION NUMBER: US 60/145,698
/ PRIOR FILING DATE: 1999-07-26
/ PRIOR APPLICATION NUMBER: US 60/146,222
/ PRIOR FILING DATE: 1999-07-28
/ PRIOR APPLICATION NUMBER: PCT/US99/20594
/ PRIOR FILING DATE: 1999-09-08
/ PRIOR APPLICATION NUMBER: PCT/US99/20944
/ PRIOR FILING DATE: 1999-09-13
/ PRIOR APPLICATION NUMBER: PCT/US99/21090
/ PRIOR FILING DATE: 1999-09-15
/ PRIOR APPLICATION NUMBER: PCT/US99/21547
/ PRIOR FILING DATE: 1999-09-15
/ PRIOR APPLICATION NUMBER: PCT/US99/23089
/ PRIOR FILING DATE: 1999-10-05
/ PRIOR APPLICATION NUMBER: PCT/US99/28214
/ PRIOR FILING DATE: 1999-11-29
/ PRIOR APPLICATION NUMBER: PCT/US99/28313
/ PRIOR FILING DATE: 1999-11-30
/ PRIOR APPLICATION NUMBER: PCT/US99/28564
/ PRIOR FILING DATE: 1999-12-02
/ PRIOR APPLICATION NUMBER: PCT/US99/28565
/ PRIOR FILING DATE: 1999-12-02
/ PRIOR APPLICATION NUMBER: PCT/US99/30095
/ PRIOR FILING DATE: 1999-12-16
/ PRIOR APPLICATION NUMBER: PCT/US99/30911
/ PRIOR FILING DATE: 1999-12-20
/ PRIOR APPLICATION NUMBER: PCT/US99/30999
/ PRIOR FILING DATE: 1999-12-20
/ PRIOR APPLICATION NUMBER: PCT/US00/00219
/ PRIOR FILING DATE: 2000-01-05
/ NUMBER OF SEQ ID NOS: 423
/ SEQ ID NO 339
/ LENGTH: 772
/ TYPE: prt
/ ORGANISM: Homo Sapien
/ US-09-902-853-339

```

[illegible]

Search completed: December 15, 2003, 17:04:34  
Job time : 31 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Comugen Ltd.

## OM protein - protein search, using SW model

Run on: December 15, 2003, 16:57:08 ; Search time 22 Seconds

(without alignments)  
251,942 Million cell updates/sec

Title: US-09-925-301-966

Perfect score: 680  
Sequence: 1 AEVHTRKQGEAPPAAMSGE.....ASNGETLEKTNRPCCVIL 131Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0  
Maximum DB seq length: 200000000Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:\*  
1: /cgn2\_6/ptodata/1/iaa/5A COMB .pep:\*  
2: /cgn2\_6/ptodata/1/iaa/5B COMB .pep:\*  
3: /cgn2\_6/ptodata/1/iaa/6A COMB .pep:\*  
4: /cgn2\_6/ptodata/1/iaa/6B COMB .pep:\*  
5: /cgn2\_6/ptodata/1/iaa/PCTUS COMB .pep:\*  
6: /cgn2\_6/ptodata/1/iaa/backfile1.ppep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysts of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	96.5	14.2	88	2	US-08-807-043-3 Sequence 3, Appl1
2	96.5	14.2	88	2	US-09-127-289-3 Sequence 3, Appl1
3	96.5	14.2	88	2	US-09-393-448-3 Sequence 3, Appl1
4	96.5	14.2	93	2	US-08-807-043-1 Sequence 1, Appl1
5	96.5	14.2	93	2	US-09-127-289-1 Sequence 1, Appl1
6	96.5	14.2	93	4	US-09-393-448-1 Sequence 1, Appl1
7	83	12.2	99	4	US-09-252-991A-17867 Sequence 17867, A
8	77	11.3	400	4	US-09-252-991A-22821 Sequence 22821, A
9	75.5	11.1	982	4	US-09-252-991A-24820 Sequence 24820, A
10	74.5	11.1	171	4	US-09-252-991A-19608 Sequence 19608, A
11	74	10.9	335	4	US-09-252-991A-26580 Sequence 26580, A
12	74	10.9	430	4	US-09-252-991A-33388 Sequence 33388, A
13	74	10.9	430	4	US-09-252-991A-20751 Sequence 20751, A
14	73.5	10.8	882	4	US-09-252-991A-17653 Sequence 17653, A
15	72.5	10.7	522	3	US-08-961-083-120 Sequence 120, App
16	72.5	10.7	522	4	US-09-536-784-120 Sequence 120, App
17	72.5	10.7	1040	3	US-08-961-083-118 Sequence 118, App
18	72.5	10.7	1040	4	US-09-536-784-118 Sequence 118, App
19	72	10.6	426	4	US-09-972-784-2 Sequence 2, Appl1
20	72	10.6	803	3	US-09-063-035-2 Sequence 2, Appl1
21	71.5	10.5	434	4	US-09-433-241A-14 Sequence 14, Appl1
22	71	10.4	333	4	US-09-562-737-1 Sequence 1, Appl1
23	70	10.3	180	4	US-09-252-991A-32768 Sequence 32768, A
24	70	10.3	571	4	US-09-252-991A-30533 Sequence 30533, A
25	69.5	10.2	344	2	US-07-857-224B-93 Sequence 93, Appl1
26	69.5	10.2	354	4	US-09-252-991A-32931 Sequence 32931, A
27	69	10.1	360	4	US-09-252-991A-28412 Sequence 28412, A

28	68.5	10.1	379	1	US-08-552-142A-11 Sequence 11, Appl1
29	68.5	10.1	381	1	US-08-910-973-11 Sequence 11, Appl1
30	68.5	10.1	381	1	US-09-499-227-11 Sequence 9, Appl1
31	68.5	10.1	382	3	US-09-234-332-9 Sequence 2, Appl1
32	68.5	10.1	409	4	US-09-252-991A-29381 Sequence 29381, A
33	68.5	10.1	878	4	US-09-735-934A-2 Sequence 2, Appl1
34	68.5	10.1	878	4	US-10-060-332-2 Sequence 2, Appl1
35	68	10.0	417	4	US-09-433-241A-10 Sequence 10, Appl1
36	68	10.0	735	3	US-09-115-704-2 Sequence 6, Appl1
37	68	10.0	1290	3	US-09-150-460B-6 Sequence 20454, A
38	67.5	9.9	376	4	US-09-252-991A-20454 Sequence 26025, A
39	67	9.9	176	4	US-09-252-991A-26025 Sequence 136, App
40	67	9.9	244	3	US-08-469-318-136 Sequence 136, App
41	67	9.9	244	3	US-08-468-609A-136 Sequence 136, App
42	67	9.9	244	4	US-08-446-872A-136 Sequence 136, App
43	67	9.9	244	4	US-08-762-227A-136 Sequence 136, App
44	67	9.9	244	5	PCT-US95-01185-136 Sequence 136, App
45	67	9.9	411	4	US-09-886-319A-31 Sequence 31, Appl1

## ALIGNMENTS

RESULT 1  
US-08-807-043-3  
; Sequence 3, Application US/08807043  
; Patent No. 5856131  
; GENERAL INFORMATION:  
; APPLICANT: Hillman, Jennifer L.  
; TITLE OF INVENTION: NOVEL HUMAN SELENOPROTEIN  
; NUMBER OF SEQUENCES: 3  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Incyte Pharmaceuticals, Inc.  
; STREET: 3174 Potter Drive  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94304  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/807,043  
; FILING DATE: Herewith  
; CLASSIFICATION: 530  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Billings, Lucy J.  
; REGISTRATION NUMBER: 36,749  
; REFERENCE/DOCKET NUMBER: PF-0202 US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 415-855-0555  
; TELEFAX: 415-845-4166  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 88 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; IMMEDIATE SOURCE:  
; LIBRARY: GenBank  
; CLONE: 993035  
; US-08-807-043-3  
Query Match 14.2%; Score 96.5; DB 2; Length 88;  
Best Local Similarity 34.1%; Pred. No. 0.0023;  
Matches 28; Conservative 17; Mismatches 26; Indels 11; Gaps 5;



```

: ATTORNEY/AGENT INFORMATION:
: NAME: Billings, Lucy J.
: REGISTRATION NUMBER: 36,749
: REFERENCE/DOCKET NUMBER: PF-0202 US
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 415-855-0555
: TELEFAX: 415-845-4166
:
: TELEX:
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 93 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: IMMEDIATE SOURCE:
: LIBRARY: BLADNOT03
: CLONE: 1599862
:
: US-09-127-289-1
:
Query Match 14.2% Score 96.5; DB 2; Length 93;
Best Local Similarity 32.8%; Pred. No. 0.0025;
Matches 21; Conservative 18; Mismatches 20; Indels 5; Gaps 3.
:
39 VRIVRVECPGPEFATYLELASAVKEQYPG-IEI--ESRLGTGAFIEINGOLVFSKLE 95
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
11 VRVV--YCGAXGVSKYLQLKKKLEDEPGRLDICGETSQAXGFEEVMVAGKLIHSCKX 68
:
QY 96 NGGF 92
:
Db 69 GDGY 72

```

RESULT 6  
 US-09-393-448-1  
 Sequence 1, Application US/09393448  
 Patent No. 6545129  
 GENERAL INFORMATION:  
 APPLICANT: Hillman, Jennifer L.  
 Goli, Surya K.  
 TITLE OF INVENTION: NOVEL HUMAN SELENOPROTEIN  
 NUMBER OF SEQUENCES: 3  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Incyte Pharmaceuticals, Inc.  
 STREET: 3174 Porter Drive  
 CITY: Palo Alto  
 STATE: CA  
 COUNTRY: USA  
 ZIP: 94304  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Diskette  
 COMPUTER: IBM Compatible  
 OPERATING SYSTEM: DOS  
 SOFTWARE: FastSeq for windows Version 2.0  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/09/393,448  
 FILING DATE: 10-Sep-1999  
 CLASSIFICATION: <Unknown>  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: 08/807,043  
 FILING DATE: <Unknown>  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Billings, Lucy J.  
 REGISTRATION NUMBER: 36,749  
 REFERENCE/DOCKET NUMBER: PP-0202 US  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 415-855-0555  
 TELEFAX: 415-845-4166  
 TELEX: <Unknown>  
 INFORMATION FOR SEQ ID NO: 1:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 93 amino acids  
 TYPE: amino acid  
 STRANDEDNESS: single

TOPOLOGY: linear  
IMMEDIATE SOURCE:  
LIBRARY: BIANOT03  
CLONE: 1599862  
SEQUENCE DESCRIPTION: SEQ ID NO: 1:  
US-09-393-448-1

Query Match 14.2%; Score 96.5; DB 4; Length 93;  
Best Local Similarity 32.8%; Pred. No. 0.0025;  
Matches 21; Conservative 18; Mismatches 20; Indels 5; Gaps 3;

QY 39 VRIYVECEPCGEATYELASAVKQYGP-IEI--ESRLGSGAFIEIINGQLVFSKLE 95  
DB 11 VRVV--YCGAXGYSKYLQTKKLEDEFGRLDICGEGTQAGXGFEVWAGKLIHKKK 68

QY 96 NGGF 99  
DB 69 GDGY 72

## RESULT 7

US-09-252-991A-17867  
Sequence 17867, Application US/09252991A  
Patent No. 6551795  
GENERAL INFORMATION:  
APPLICANT: Marc J. Rubenfield et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
FILE REFERENCE: 107196.136  
CURRENT APPLICATION NUMBER: US/09/252,991A  
PRIOR FILING DATE: 1999-02-18  
PRIOR APPLICATION NUMBER: US 60/074,788  
PRIOR FILING DATE: 1998-02-18  
PRIOR APPLICATION NUMBER: US 60/094,190  
PRIOR FILING DATE: 1998-07-27  
NUMBER OF SEQ ID NOS: 33142  
SEQ ID NO 17867  
LENGTH: 99  
TYPE: PRT  
ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-17867

Query Match 12.2%; Score 83; DB 4; Length 99;  
Best Local Similarity 31.2%; Pred. No. 0.083;  
Matches 22; Conservative 13; Mismatches 36; Indels 6; Gaps 3;

QY 35 PGSGVRIYVECEPCG--EATYI--ELASAVKQYGCIEIESRLGSGAFIEIINGQLV 90  
DB 5 PTAKEPIVITYCTQCWLTPAAWLAQELSTFADLDGKVCLEPGTG--GVFRITCDGVQV 62

QY 91 FSKLENGGFPEYKDLIAIR 110  
DB 63 WERKADGFPPEAKLKORVR 82

## RESULT 8

US-09-252-991A-22821  
Sequence 22821, Application US/09252991A  
Patent No. 6551795  
GENERAL INFORMATION:  
APPLICANT: Marc J. Rubenfield et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
FILE REFERENCE: 107196.136  
CURRENT APPLICATION NUMBER: US/09/252,991A  
PRIOR FILING DATE: 1999-02-18  
PRIOR APPLICATION NUMBER: US 60/074,788  
PRIOR FILING DATE: 1998-02-18  
PRIOR APPLICATION NUMBER: US 60/094,190  
PRIOR FILING DATE: 1998-07-27  
NUMBER OF SEQ ID NOS: 33142  
SEQ ID NO 22821  
LENGTH: 400

TYPE: PRT  
ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-22821

Query Match 11.3%; Score 77; DB 4; Length 400;  
Best Local Similarity 27.5%; Pred. No. 2.7;  
Matches 30; Conservative 12; Mismatches 41; Indels 26; Gaps 5;

QY 7 KCGPEAPEAMSGEGQTSVAP-----PPEVEP-----GSGVRIYVECEPCG 50  
DB 276 QQPTVTAAPGATPAAGTAPAVPATASPAAPPAPEPAAAPVAGGGQV-VKVGFAVDCW 334

QY 51 FEATYLE--LASAVKQYGPGE-----IESRLGSGAFIEIINGQLV 90  
DB 335 TQVTDANGKVLVSALKRKGDSELAGKAPLEIRLDFARGAQVSYNGQPV 383

## RESULT 9

US-09-252-991A-24820  
Sequence 24820, Application US/09252991A  
Patent No. 6551795  
GENERAL INFORMATION:  
APPLICANT: Marc J. Rubenfield et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
FILE REFERENCE: 107196.136  
CURRENT APPLICATION NUMBER: US/09/252,991A  
PRIOR FILING DATE: 1999-02-18  
PRIOR APPLICATION NUMBER: US 60/074,788  
PRIOR FILING DATE: 1998-02-18  
PRIOR APPLICATION NUMBER: US 60/094,190  
PRIOR FILING DATE: 1998-07-27  
NUMBER OF SEQ ID NOS: 33142  
SEQ ID NO 24820  
LENGTH: 982  
TYPE: PRT  
ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-24820

Query Match 11.1%; Score 75.5; DB 4; Length 982;  
Best Local Similarity 27.7%; Pred. No. 14;  
Matches 26; Conservative 12; Mismatches 23; Indels 33; Gaps 4;

QY 4 HTRKQPEAPEAMSGEPQTSVAPPEVEPESGSGVRIYVECEPCGEATYELASAVK 63  
DB 333 HGEATTPAPAPALGSGGKT---PWEPIAP-----ELASLV- 367

QY 64 EYVPGIEIESRLGSGAFIEIINGQLVFSKLENG 97  
DB 368 EQPP-----RGEWAYEIKLDGYRLMSRIEDG 393

## RESULT 10

US-09-252-991A-19608  
Sequence 19608, Application US/09252991A  
Patent No. 6551795  
GENERAL INFORMATION:  
APPLICANT: Marc J. Rubenfield et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
FILE REFERENCE: 107196.136  
CURRENT APPLICATION NUMBER: US/09/252,991A  
PRIOR FILING DATE: 1999-02-18  
PRIOR APPLICATION NUMBER: US 60/074,788  
PRIOR FILING DATE: 1998-02-18  
PRIOR APPLICATION NUMBER: US 60/094,190  
PRIOR FILING DATE: 1998-07-27  
NUMBER OF SEQ ID NOS: 33142  
SEQ ID NO 19608  
LENGTH: 171  
TYPE: PRT  
ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-19608



Query Match 11.0%; Score 74.5; DB 4; Length 171;  
Best Local Similarity 27.8%; Pred. No. 1.6;  
Matches 20; Conservative 11; Mismatches 38; Indels 3; Gaps 1;

QY 6 RKQGEPEAPAMGEPGQTSVAPPPEVEPGS---GVRIVCEPCGFEATYIELASAV 62  
DB 20 RASGSSSPRTSSQPSGRSMIOCKRYEPASPADGRVLDLWPRGIRKALAMDWL 79  
QY 63 KEQYPGIEISR 74  
DB 80 KDVAFNSNEVAKQ 91

RESULT 11  
US-09-252-991A-26580  
; Sequence 26580, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; PRIOR FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 26580  
; LENGTH: 325  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-26580

Query Match 10.9%; Score 74; DB 4; Length 325;  
Best Local Similarity 36.7%; Pred. No. 4.4;  
Matches 18; Conservative 8; Mismatches 9; Indels 14; Gaps 2;

QY 6 RKQG-----PEAPEAMGEPGQTSVAPP-----PEVEPGSGVR 40  
DB 230 RRRGSLPVGRRPARPAVARGPGRTOAVPAQPRLPGRPAQDPGAGLR 278

RESULT 12  
US-09-252-991A-31388  
; Sequence 31388, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; PRIOR FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 31388  
; LENGTH: 430  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-31388

Query Match 10.9%; Score 74; DB 4; Length 430;  
Best Local Similarity 27.2%; Pred. No. 6.5;  
Matches 40; Conservative 16; Mismatches 43; Indels 48; Gaps 10;

QY 6 RKQG---PEAPEAMG-----GEPGQTSVAPPPEVEPGSGVRIVCEYK 47  
DB 63 KEQYPGIEISR 74  
DB 80 KDVAFNSNEVAKQ 91

DB 244 RQGVFLPAAAPQAVAAQEDLOVEDALORRRGDPYRRAVA-DEAQVEAGEA-VNEAGLR 301  
QY 48 PCGEATYLEIASAVEQYPGIEISRLGSGTGAFEIENGQVLSKLENGCFPEYKD--- 104  
DB 302 PAGGRAY-----AVTDPV-VDIAGRIGVGTFEHPA-----ALAEQGI-HGDDTQA 348  
QY 105 ---LIEAIRRANGSETLEKITSRPP 127  
DB 349 AIVVVIARRQAPGAE-----RPP 367

RESULT 13  
US-09-252-991A-20751  
; Sequence 20751, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; PRIOR FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 20751  
; LENGTH: 492  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-20751

Query Match 10.9%; Score 74; DB 4; Length 492;  
Best Local Similarity 39.6%; Pred. No. 7.8;  
Matches 21; Conservative 6; Mismatches 24; Indels 2; Gaps 2;

QY 1 AEVHTKQGEPEAPAMGEPGQ-TSVAPPPEVEPGSGVRIVCEYK 52  
DB 7 ASARTRLRPPAMPRRSSARPEPTDGARSPPRLP-SGPRVLVLAHFGIG 58

RESULT 14  
US-09-252-991A-17653  
; Sequence 17653, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; PRIOR FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 17653  
; LENGTH: 882  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
; NAME/KEY: UNSURE  
; LOCATION: (730)  
; OTHER INFORMATION: Identity of amino acid at the above locations are unknown.  
US-09-252-991A-17653

Query Match 10.8%; Score 73.5; DB 4; Length 882;  
Best Local Similarity 24.5%; Pred. No. 20;  
Matches 24; Conservative 14; Mismatches 37; Indels 23; Gaps 3;

QY 3 VHTRKQGEPEAPAMGEPGQTSVAPPPEVEPGSGVRIVCEYK 53  
DB 63 KEQYPGIEISR 74  
DB 80 KDVAFNSNEVAKQ 91

Db 143 VDAEOEADRMVPAFAVAALGTLPAGARLEPAFG-----SLPAGLAVAGLAAQAFAD 195  
QY 54 TYVELASAVEQO-----YPGIEIESRLGTGAFEIE 84  
Db 196 AVAQIAAVEQOAGERRVDFPGQLLEBPLGCGVFRQE 233

## RESULT 15

US-08-961-083-120  
; Sequence 120, Application US/08961083  
; Patent No. 6159469  
; GENERAL INFORMATION:  
; APPLICANT: Choi et. al.  
; TITLE OF INVENTION: Streptococcus pneumoniae Antigen and Vaccines  
; NUMBER OF SEQUENCES: 452  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Human Genome Sciences, Inc.  
; STREET: 9410 Key West Avenue  
; CITY: Rockville  
; STATE: Maryland  
; COUNTRY: USA  
; ZIP: 20850  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4MB storage  
; COMPUTER: HP Vectra 486/33  
; OPERATING SYSTEM: MSDOS version 6.2  
; SOFTWARE: ASCII Text  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/961,083  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Brookes, A. Anders  
; REGISTRATION NUMBER: 36,373  
; REFERENCE/DOCKET NUMBER: PB340P2  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (301) 309-8504  
; TELEFAX: (301) 309-8512  
; INFORMATION FOR SEQ ID NO: 120:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 522 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-961-083-120

Query March 10.7%; Score 72.5; DB 3; Length 522;

Best Local Similarity 23.4%; Pred. No. 12;  
Matches 29; Conservative 14; Mismatches 56; Indels 25; Gaps 3;

QY 5 TRKQPPAEPAPMAGSGPQTSVAPPPEVEBPGSVRIIV-----EYCEPCGFATY--L 56  
Db 250 TQGGHGEAARREBPATYPLATKQPGHKGKATVREBTLETPVATKGTQEPBH 309  
QY 57 ELASAVEQYPGIEIESRLGTGAFEIEINGQLVFSKLKNGGFPYKDLIEAIRASNGE 116  
Db 310 EGRXVVEELPALAVTTR-----NRTIIONIPYTTEIIOPTLLKNRR 352  
QY 117 TLER 120  
Db 353 KIER 356

Search completed: December 15, 2003, 17:00:15  
Job time : 29 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 15, 2003, 17:03:56 ; Search time 21 Seconds  
(without alignments)  
599.909 Million cell updates/sec

Title: US-09-925-301-966

Perfect score: 131

Sequence: 1 AEVHTRKQGPPEAPPAAMSGE.....ASNGETLEKITNSRPPCVIL 131

Scoring table: OLIGO

Searched: 283308 seqs, 96168682 residues

Word size : 0

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 100 summaries

Database :

1: p1r76:\*  
2: p1r1:\*  
3: p1r3:\*  
4: p1r4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	9	6.9	158	2	S36561
2	8	6.1	220	2	D64316
3	7	5.3	78	2	AB1554
4	7	5.3	88	2	S68159
5	7	5.3	160	2	B70718
6	7	5.3	217	2	AG2623
7	7	5.3	217	2	F97405
8	7	5.3	245	2	T10130
9	7	5.3	258	2	T46491
10	7	5.3	260	2	A36949
11	7	5.3	280	2	A69899
12	7	5.3	307	2	B81196
13	7	5.3	373	2	H84404
14	7	5.3	375	2	T03681
15	7	5.3	382	2	I19068
16	7	5.3	421	2	A99409
17	7	5.3	445	2	C64172
18	7	5.3	454	2	AG1977
19	7	5.3	467	2	D72461
20	7	5.3	473	2	T28052
21	7	5.3	483	2	AF3373
22	7	5.3	488	2	T36098
23	7	5.3	494	2	B84266
24	7	5.3	548	2	B90446
25	7	5.3	555	2	T48453
26	7	5.3	601	2	T37738
27	7	5.3	705	2	A70669
28	7	5.3	716	2	C53292
29	7	5.3	820	2	D71471

30	7	5.3	980	2	A38523	genome polyprotein
31	7	5.3	1080	2	T19048	probable Pro-X car
32	7	5.3	1170	2	A57650	repar protein Xpg
33	6	4.6	22	2	P00667	phocystem I 19.3
34	6	4.6	30	2	B97906	phospho-beta-gluc
35	6	4.6	45	1	D69466	conserved hypotet
36	6	4.6	52	1	A40424	phospholamban - hu
37	6	4.6	52	1	S37638	phospholamban - ra
38	6	4.6	52	1	A49057	phospholamban - mo
39	6	4.6	52	1	B40424	phospholamban - ra
40	6	4.6	52	1	A29002	phospholamban - do
41	6	4.6	52	1	S05540	phospholamban - pi
42	6	4.6	56	2	AF2153	hypothetical prote
43	6	4.6	62	2	AH2604	conserved hypotet
44	6	4.6	62	2	G97386	hypothetical 12.7K
45	6	4.6	64	2	JH0338	T-cell receptor al
46	6	4.6	64	2	A13203	conserved hypotet
47	6	4.6	74	2	C42266	peptidylglycine mo
48	6	4.6	75	2	D87654	transcription regu
49	6	4.6	75	2	A11634	protein encoded by
50	6	4.6	76	2	A43537	heat-stable anti
51	6	4.6	76	2	I53107	CD24 precursor - r
52	6	4.6	77	2	C95014	conserved hypotet
53	6	4.6	77	2	F97887	conserved hypotet
54	6	4.6	82	2	C75296	transcription regu
55	6	4.6	83	2	S75314	phycocyanin-assoc
56	6	4.6	86	2	E86435	protein F17F8.11 l
57	6	4.6	89	2	F75397	hypothetical prote
58	6	4.6	91	2	T08131	oleosin-like prote
59	6	4.6	92	2	T08129	related to spore c
60	6	4.6	93	2	C97230	probable membrane
61	6	4.6	96	2	T35372	T-cell receptor al
62	6	4.6	99	2	JH0337	hypothetical prote
63	6	4.6	100	2	AP1490	probable membrane
64	6	4.6	102	2	S51974	hypothetical prote
65	6	4.6	103	2	E81030	hypothetical prote
66	6	4.6	103	2	C81061	hypothetical prote
67	6	4.6	104	2	AC1130	hypothetical prote
68	6	4.6	105	2	JH0083	T-cell receptor de
69	6	4.6	105	2	T13328	hypothetical prote
70	6	4.6	106	2	S03476	T-cell receptor al
71	6	4.6	107	2	D84391	hypothetical prote
72	6	4.6	107	2	A70966	hypothetical prote
73	6	4.6	109	2	JH0335	T-cell receptor al
74	6	4.6	110	2	S23896	T-cell receptor al
75	6	4.6	110	2	S23368	probable S6 riboso
76	6	4.6	112	2	D81732	genome polyprotein
77	6	4.6	112	2	S51124	T-cell receptor al
78	6	4.6	113	2	A29281	T-cell receptor al
79	6	4.6	114	2	S07703	probable membrane
80	6	4.6	114	2	S69320	hypothetical prote
81	6	4.6	114	2	D90784	hypothetical prote
82	6	4.6	114	2	B90903	hypothetical prote
83	6	4.6	114	2	B85644	hypothetical prote
84	6	4.6	114	2	B85713	unknown protein en
85	6	4.6	115	2	A46242	T-cell receptor al
86	6	4.6	115	2	S70185	hypothetical prote
87	6	4.6	116	1	A48344	trans-activating t
88	6	4.6	116	1	H82478	hypothetical 13.2K
89	6	4.6	118	2	D32534	T cell receptor CK
90	6	4.6	120	2	S57887	ribosomal protein
91	6	4.6	120	2	T44399	T cell receptor Bb
92	6	4.6	122	2	S57888	T-cell receptor al
93	6	4.6	125	1	A31557	cytochrome c' - Rh
94	6	4.6	126	1	CCOFCR	T-cell receptor al
95	6	4.6	126	2	S03479	aspartate 1-decarb
96	6	4.6	127	2	AD1312	aspartate 1-decarb
97	6	4.6	127	2	AD1684	T-cell receptor al
98	6	4.6	128	2	E24402	MHC class II I-A-a
99	6	4.6	128	2	I71933	
100	6	4.6	132	2	I71933	

## ALIGNMENTS

## RESULT 1

S36561  
E6 protein - human papillomavirus type 45  
C/Species: human papillomavirus type 45  
C/Date: 20-Feb-1995 #sequence\_revision 20-Feb-1995 #text\_change 26-Aug-1999  
C/Accession: S36561  
R/Deletion: H.; Hofmann, B.  
submitted to the EMBL Data Library, August 1993  
A/Description: Primer-directed sequencing of human papillomavirus types.  
A/Reference number: S36469  
A/Accession: S36561  
A/Molecule type: DNA  
A/Residues: 1-158 <DEL>  
A/Cross-references: EMBL:X74479; NID:G397022; PIDN:CA52573.1; PID:G397023  
C/Superfamily: papillomavirus E6 protein  
C/Keywords: DNA binding; early protein; nucleus; zinc finger

Query Match  
Best Local Similarity 100.0%; Pred. No. 0.2;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 115 GETLEKTN 123  
DB 87 GETLEKTN 95

## RESULT 2

D64316  
restriction modification enzyme subunit M1 homolog - Methanococcus jannaschii  
C/Species: Methanococcus jannaschii  
C/Date: 13-Sep-1996 #sequence\_revision 13-Sep-1996 #text\_change 16-Aug-2002  
C/Accession: D64316  
R/Built: C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blake,  
; Reich, C.T.; Overbeek, R.; Kirkness, E.F.; Weinstock, K.G.; Merrick, J.M.; Glodek, A.;  
rson, J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hurt, M.A.  
Science 273, 1058-1073, 1996  
A/Authors: Kahne, B.P.; Borodovsky, M.; Klenk, H.P.; Frazer, C.M.; Smith, H.O.; Moese,  
A/Title: Complete genome sequence of the methanogenic archaeon, Methanococcus jannaschii  
A/Reference number: A64300; MUID:96337999; PMID:8688087  
A/Accession: D64316  
A/Status: preliminary; nucleic acid sequence not shown; translation not shown  
A/Molecule type: DNA  
A/Residues: 1-220 <BDL>  
A/Cross-references: GB:U67470; GB:U77117; NID:G3826247; PIDN:AAB98113.1; PID:G1592267; T  
C/Genetics:  
A/Map position: REV127472-126810  
A/Start codon: TTG  
C/Superfamily: type I site-specific deoxyribonuclease chain hsdM (associate member)

Query Match  
Best Local Similarity 100.0%; Pred. No. 2.7;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 102 EKDLIEAI 109  
DB 73 EKDLIEAI 80

## RESULT 3

AB1554  
D-alanyl carrier protein [imported] - Listeria innocua (strain C1p11262)  
C/Species: Listeria innocua  
C/Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 27-Nov-2001  
C/Accession: AB1554  
R/Glaeser, P.; Franke, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bioecker  
; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurge, O.; Entian, K.D.; Fsihl, H.  
Science 294, 849-852, 2001  
A/Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maltournam, A.; Ma  
ok, C.; Schluter, T.; Simoes, N.; Tietzer, A.; Vazquez-Boland, J.A.; Voss, H.; Wehlend,

A/Title: Comparative genomics of Listeria species.

A/Reference number: AB1077; MUID:21537279; PMID:11679669  
A/Accession: AB1554  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-78 <GLA>  
A/Cross-references: GB:AU592022; PIDN:CAC96202.1; PID:G16413430; GSPDB:GN00178  
A/Experimental source: strain C1p11262  
C/Genetics:  
A/Gene: dltC

Query Match  
Best Local Similarity 100.0%; Pred. No. 11;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 69 IEIESRL 75  
DB 44 IEIESRL 50

## RESULT 4

S68159  
H+-transporting two-sector ATPase (EC 3.6.3.14) chain 9 - slime mold (Dictyostelium disc  
C/Species: Dictyostelium discoidium  
C/Date: 06-Dec-1996 #sequence\_revision 13-Mar-1997 #text\_change 03-Jun-2002  
C/Accession: S68159; T43761  
R/Angata, K.; Kuroe, K.; Yanagisawa, K.; Tanaka, Y.  
Curr. Genet. 27, 249-256, 1995  
A/Title: Codon usage, genetic code and phylogeny of Dictyostelium discoidium mitochondria  
A/Reference number: S68155; MUID:9525468; PMID:7736610  
A/Accession: S68159  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-88 <ANG>  
A/Cross-references: EMBL:D16466; NID:G695950; PIDN:BA03937.1; PID:G695955  
R/Gawa, S.; Yoshino, R.; Angata, K.; Pi, M.; Iwamoto, M.; Kuroe, K.; Matsuo, K.; Morio,  
submitted to the EMBL Data Library, December 1996  
A/Description: The mitochondrial DNA of Dictyostelium discoidium. Complete sequence, gene  
A/Reference number: Z22666  
A/Accession: T43761  
A/Status: preliminary; translated from GB/EMBL/DDBL  
A/Molecule type: DNA  
A/Residues: 1-88 <OGA>  
A/Cross-references: EMBL:AB000109; PIDN:BAA78065.1  
C/Genetics:  
A/Gene: ATP9  
A/Genome: mitochondrion  
C/Superfamily: H+-transporting ATP synthase lipid-binding protein  
C/Keywords: hydrolase; mitochondrion

Query Match  
Best Local Similarity 100.0%; Pred. No. 13;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 56 LELASAV 62  
DB 10 LELASAV 16

RESULT 5  
B70718  
hypothetical protein RV0964c - Mycobacterium tuberculosis (strain H37RV)

C/Species: Mycobacterium tuberculosis  
C/Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 20-Jun-2000  
C/Accession: B70718  
R/Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S  
; Connor, R.; Davies, R.; Devlin, K.; Feltham, T.; Gentles, S.; Hamlin, N.; Holtroyd, S.;  
Rajandream, M.A.; Rogers, J.; Rutter, K.; Seeger, K.; Skelton, S.; Squares, S.  
Nature 393, 537-544, 1998  
A/Authors: Sgares, R.; Sultson, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
A/Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome  
A/Reference number: A70500; MUID:98295987; PMID:9634230  
A/Accession: B70718

A>Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A/Molecule type: DNA  
 A/Residues: 1-160 <COL>  
 A/Cross-references: GB:279700; GB:AL123456; NID:g3261628; PIDN:CAB01987.1; PID:g1524198  
 A/Experimental source: strain H378v  
 C/Genetics:  
 A/Gene: RV0964c  
 C/Superfamily: Mycobacterium tuberculosis hypothetical protein RV0964c

Query Match 5.3%; Score 7; DB 2; Length 160;  
 Best Local Similarity 100.0%; Pred. No. 21;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 108 AIRRASN 114  
 DB 95 AIRRASN 101

RESULT 6  
 AG2623  
 GRP-binding protein Atu0383 [imported] - Agrobacterium tumefaciens (strain C58, Dupont)  
 C/Species: Agrobacterium tumefaciens  
 C/Date: 11-Jan-2002 #sequence\_revision 11-Jan-2002 #text\_change 18-Nov-2002  
 C/Accession: AG2623  
 R/Wood, D.W.; Secubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, I.  
 erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutyavlin, T.; Levy, R.; Li, M.; McClellan,  
 i; Karp, P.; Romero, P.; Zhang, S.  
 Science 294, 2317-2323, 2001  
 A/authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,  
 ster, E.W.  
 A/Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.  
 A/Reference number: AB2577; MUID:21608550; PMID:11743193  
 A/Accession: AG2623  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-217 <KUR>  
 A/Cross-references: GB:AE008688; PIDN:AAL41405.1; PID:g17738724; GSPDB:SN00186  
 A/Experimental source: strain C58 (Dupont)  
 C/Genetics:  
 A/Gene: Atu0383  
 A/Map position: circular chromosome  
 C/Superfamily: Bacillus subtilis conserved hypothetical protein yxXC; translation elonga

Query Match 5.3%; Score 7; DB 2; Length 217;  
 Best Local Similarity 100.0%; Pred. No. 28;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 115 GETLEKI 121  
 DB 177 GETLEKI 183

RESULT 7  
 P97405  
 GRP-binding protein cgpA (AF180145) [imported] - Agrobacterium tumefaciens (strain C58,  
 C/Species: Agrobacterium tumefaciens  
 C/Date: 30-Sep-2001 #sequence\_revision 30-Sep-2001 #text\_change 18-Nov-2002  
 C/Accession: P97405  
 R/Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman,  
 A.; Liu, F.; Woliam, C.; Allinger, M.; Doughy, D.; Scott, C.; Lappae, C.; Markelz, B.;  
 Science 294, 2323-2328, 2001  
 A/Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum  
 A/Reference number: A97359; MUID:21608551; PMID:11743194.  
 A/Accession: P97405  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-217 <KUR>  
 A/Cross-references: GB:AE007869; PIDN:AAK86199.1; PID:g15155296; GSPDB:SN00169  
 C/Genetics:  
 A/Gene: AGR\_C\_672  
 A/Map position: circular chromosome  
 C/Superfamily: Bacillus subtilis conserved hypothetical protein yxXC; translation elonga

Query Match 5.3%; Score 7; DB 2; Length 217;  
 Best Local Similarity 100.0%; Pred. No. 28;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 115 GETLEKI 121  
 DB 177 GETLEKI 183

RESULT 8  
 T10130  
 hypothetical protein - phage T4  
 C/Species: phage T4  
 C/Date: 16-Jul-1999 #sequence\_revision 16-Jul-1999 #text\_change 16-Jul-1999  
 C/Accession: T10130  
 R/Chu, F.  
 submitted to the EMBL Data Library, April 1998  
 A/Reference number: Z16963  
 A/Accession: T10130  
 A/Status: preliminary; translated from GB/EMBL/DBJ  
 A/Molecule type: DNA  
 A/Residues: 1-245 <CHU>  
 A/Cross-references: EMBL:M12742; NID:g3033366; PID:g3033368  
 A/Experimental source: strain alc4

Query Match 5.3%; Score 7; DB 2; Length 245;  
 Best Local Similarity 100.0%; Pred. No. 31;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 100 PYEKDLI 106  
 DB 65 PYEKDLI 71

RESULT 9  
 T46491  
 hypothetical protein DKFZp434P086.1 - human  
 C/Species: Homo sapiens (man)  
 C/Date: 04-Feb-2000 #sequence\_revision 04-Feb-2000 #text\_change 04-Feb-2000  
 C/Accession: T46491  
 R/Koerner, K.; Beyer, A.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.  
 submitted to the Protein Sequence Database, January 2000  
 A/Reference number: Z23037  
 A/Accession: T46491  
 A/Status: preliminary  
 A/Molecule type: mRNA  
 A/Residues: 1-258 <AAA>  
 A/Cross-references: EMBL:AL137662  
 A/Experimental source: adult testis; clone DKFZp434P086  
 C/Genetics:  
 A/Note: DKFZp434P086.1

Query Match 5.3%; Score 7; DB 2; Length 258;  
 Best Local Similarity 100.0%; Pred. No. 33;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 27 APPPEEV 33  
 DB 155 APPPEEV 161

RESULT 10  
 A36949  
 28.9K basic DNA-binding protein - Mycobacterium scrofulaceum plasmid pMSG262  
 C/Species: Mycobacterium scrofulaceum  
 C/Date: 06-Oct-1994 #sequence\_revision 25-Apr-1997 #text\_change 25-Apr-1997  
 C/Accession: A36949  
 R/Oin, M.; Taniguchi, H.; Mizuguchi, Y.  
 J. Bacteriol. 176, 419-425, 1994  
 A/Title: Analysis of the replication region of a mycobacterial plasmid, pMSG262.  
 A/Reference number: A36949; MUID:94117377; PMID:8288537  
 A/Accession: A36949  
 A/Status: preliminary

A:Molecule type: DNA  
 A:Residues: 1-260 <QIN>  
 A:Name: sequence extracted from NCBI backbone (NCBIN:142814, NCBI:142815)  
 A:Genetics:  
 A:Genome: plasmid  
 C:Keywords: DNA binding

Query Match	Similarity	Score	DB	Length
Best Local	100.0%	Pred. No. 33		
Matches	7	Conservative	0	Mismatches
			0	Indels
			0	Gaps
			0	

OY 108 AIRRASN 114  
 DB 178 AIRRASN 184

## RESULT 11

A69899  
 hypothetical protein yobJ - Bacillus subtilis

C:Species: Bacillus subtilis  
 C>Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 28-Jul-2000  
 C:Accession: A69899  
 R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Bertozzi, C.; Bron, S.; Brouillet, S.; Bruch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Cho, A.; Ehrlich, S.D.; Emerson, P.T.; Ertlan, K.D.; Errington, J.; Fabret, C.; Ferrari, E.; Nature 390, 249-256, 1997  
 A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gallier, A.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.; Koester, P.; Konings, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois, T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K.; Authors: Yoshikawa, H.F.; Zumschein, E.; Yoshikawa, H.; Danchin, A.  
 A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.  
 A:Reference number: A69880; PMID:98044033; PMID:9384377  
 A:Accession: A69899  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-280 <RIN>  
 A:Cross-references: GB:299114; GB:AL009126; NID:g2634230; PIDN:CAB13790.1; PID:el165370;  
 A:Experimental source: strain 168  
 C:Genetics:  
 A:Gene: yobJ  
 C:Superfamily: Bacillus subtilis hypothetical protein yobJ

Query Match	Similarity	Score	DB	Length
Best Local	100.0%	Pred. No. 35		
Matches	7	Conservative	0	Mismatches
			0	Indels
			0	Gaps
			0	

OY 115 GETLEKI 121  
 DB 43 GETLEKI 49

## RESULT 12

B81196  
 agmatinase (EC 3.5.3.11) NMA2016 [imported] - Neisseria meningitidis (strain MC58 serogroup C:Species: Neisseria meningitidis

C>Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 02-Feb-2001  
 C:Accession: B81196; C81831  
 R:Teitelin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, J.A.; Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.A.; Sci, H.; Olt, H.; Vamthyan, J.; Gill, J.; Scarlato, V.; Massignani, V.; Pizza, M.  
 A:Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58.  
 A:Reference number: A81000; PMID:20175755; PMID:10710307  
 A:Accession: B81196  
 A:Status: preliminary  
 A:Molecule type: DNA

A:Residues: 1-307 <TER>  
 A:Cross-references: GB:AE002403; GB:AE002098; NID:g7225688; PIDN:AAE40906.1; PID:g7225696  
 A:Experimental source: serogroup B, strain MC58  
 R:Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Morel, J.; Holt, S.; Jagers, K.; Leather, S.; Moul, S.; Mungall, K.; Quail, M.A.; Rajandream, Nature 404, 502-506, 2000  
 A:Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis 22491.  
 A:Reference number: A81775; PMID:20225566; PMID:10761919  
 A:Accession: C81831  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-307 <PAR>  
 A:Cross-references: GB:AL162757; GB:AL157959; NID:g7380371; PIDN:CAB85255.1; PID:g738064;  
 A:Experimental source: serogroup A, strain 22491  
 C:Genetics:  
 A:Gene: speB; NMB0469; NMA2016  
 C:Superfamily: arginase  
 C:Keywords: hydrolase

Query Match	Similarity	Score	DB	Length
Best Local	100.0%	Pred. No. 38		
Matches	7	Conservative	0	Mismatches
			0	Indels
			0	Gaps
			0	

OY 107 EAIRRAS 113  
 DB 59 EAIRRAS 65

## RESULT 13

H84404  
 ferrichrome ABC transporter permease [imported] - Halobacterium sp. NRC-1

C:Species: Halobacterium sp. NRC-1  
 C>Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 17-May-2002  
 C:Accession: H84404  
 R:Ng, W.V.; Kennedy, S.P.; Mahairas, G.G.; Bergquist, B.; Pan, M.; Shukla, H.D.; Laeky, S.; Leihauer, B.; Keller, K.; Cruz, R.; Danison, M.J.; Hough, D.W.; Maddocks, D.G.; Jablor, Jung, K.H.; Alam, M.; Freitas, T.  
 A:Title: Genome sequence of Halobacterium species NRC-1.  
 A:Reference number: A84160; PMID:20504483; PMID:11016950  
 A:Accession: H84404  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-373 <STO>  
 A:Cross-references: GB:AE004437; NID:g10581939; PIDN:AA20604.1; GSPDB:GN00138  
 C:Genetics:  
 A:Gene: fnuG  
 C:Superfamily: vitamin B12 transport protein btuc

Query Match	Similarity	Score	DB	Length
Best Local	100.0%	Pred. No. 45		
Matches	7	Conservative	0	Mismatches
			0	Indels
			0	Gaps
			0	

OY 75 LGGTGAF 81  
 DB 163 LGGTGAF 169

## RESULT 14

T03681  
 putrescine N-methyltransferase (EC 2.1.1.53) A411 [validated] - common tobacco

C:Species: Nicotiana tabacum (common tobacco)  
 C>Date: 24-Mar-1999 #sequence\_revision 24-Mar-1999 #text\_change 21-Jul-2000  
 C:Accession: T03681  
 R:Hibi, N.; Higashiguchi, S.; Hashimoto, T.; Yamada, Y.  
 A:Title: Gene expression in tobacco low-nicotine mutants.  
 A:Reference number: Z14614; PMID:94312878; PMID:8038607  
 A:Accession: T03681  
 A:Status: preliminary; translated from GB/EMBL/DBCB  
 A:Molecule type: mRNA  
 A:Residues: 1-375 <HIB>

A;Cross-references: EMBL:D28506; NID:g507824; PIDN:BA05867.1; PID:g507825  
 A;Experimental source: strain Burley 2; tissue-type root  
 C;Superfamily: spermidine synthase  
 C;Keywords: methyltransferase; S-adenosylmethionine

Query Match 5.3%; Score 7; DB 2; Length 375;  
 Best Local Similarity 100.0%; Pred. No. 45;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 95 ENGPPY 101  
 DB 141 ENGPPY 147

RESULT 15  
 139068  
 proline- arginine-rich end leucine-rich repeat protein PRELP precursor - human  
 C;Species: Homo sapiens (man)  
 C;Date: 01-Mar-1996 #sequence\_revision 01-Mar-1996 #text\_change 24-Sep-1999  
 C;Accession: I39068  
 R;Benson, E.; Neame, P.J.; Helnegard, D.; Sommerlin, Y.  
 J. Biol. Chem. 270, 25639-25644, 1995  
 A;Title: The primary structure of a basic leucine-rich repeat protein, PRELP, found in  
 A;Reference number: I39068; MUID:96029653; PMID:7592739  
 A;Accession: I39068  
 A;Status: preliminary  
 A;Molecule type: mRNA  
 A;Residues: 1-382 <RES>  
 A;Cross-references: EMBL:U29089; NID:g886135; PIDN:AAC50230.1; PID:g886136  
 C;Genetics:  
 A;Gene: GDB:PRELP  
 A;Cross-references: GDB:696218  
 A;Map position: 1q32.1-1q32.1  
 C;Superfamily: fibromodulin; leucine-rich alpha-2-glycoprotein repeat homology  
 C;Keywords: tandem repeat

Query Match 5.3%; Score 7; DB 2; Length 382;  
 Best Local Similarity 100.0%; Pred. No. 46;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 90 VPSKLEN 96  
 DB 190 VPSKLEN 196

RESULT 16  
 A99409  
 conserved hypothetical protein [imported] - Sulfolobus solfataricus  
 C;Species: Sulfolobus solfataricus  
 C;Date: 24-May-2001 #sequence\_revision 24-May-2001 #text\_change 15-Jun-2001  
 C;Accession: A99409  
 R;She, Q.; Singh, R.K.; Confalonieri, F.; Zivanovic, Y.; Allard, G.; Ayyez, M.J.; Chan-  
 Jong, I.; Ueffing, A.C.; Kozera, C.J.; Medina, N.; Peng, X.; Thi-Ngoc, H.P.; Redder, R.  
 et al. R.A.; Ragan, M.A.; Jensen, C.W.; Van der Oost, J.  
 submitted to Genbank, April 2001  
 A;Description: Sulfolobus solfataricus complete genome.  
 A;Reference number: A99139  
 A;Accession: A99409  
 A;Status: preliminary  
 A;Molecule type: DNA  
 A;Residues: 1-421 <NUR>  
 A;Cross-references: GB:AE006641; NID:g13815682; PIDN:AAK42528.1; GSPDB:GN00155  
 C;Genetics:  
 A;Gene: SSO2380  
 C;Superfamily: conserved hypothetical protein b0835

Query Match 5.3%; Score 7; DB 2; Length 421;  
 Best Local Similarity 100.0%; Pred. No. 50;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 104 DLIEAR 110  
 DB 232 DLIEAR 238

RESULT 17  
 C64172  
 hypothetical protein H11463 - Haemophilus influenzae (strain Rd KW20)  
 N;Alternate names: hypothetical protein H11337  
 C;Species: Haemophilus influenzae  
 C;Date: 18-Aug-1995 #sequence\_revision 18-Aug-1995 #text\_change 17-Mar-2000  
 C;Accession: C64172; A64171  
 R;Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage, A.  
 ; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman, J.  
 , D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geoghegan, N.S.M.  
 Science 269, 496-512, 1995  
 A;Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter, J.  
 A;Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.  
 A;Reference number: A64000; MUID:95505630; PMID:7542800

A;Accession: C64172  
 A;Status: nucleic acid sequence not shown; translation not shown  
 A;Molecule type: DNA  
 A;Residues: 1-445 <TRIG>  
 A;Cross-references: GB:U32824; GB:I42023; NID:g1574298; PIDN:AAC23110.1; PID:g1574303; T1  
 A;Note: best homolog was a hypothetical protein from Escherichia coli  
 A;Accession: A64171  
 A;Status: nucleic acid sequence not shown; translation not shown  
 A;Molecule type: DNA  
 A;Residues: 1-445 <TRIG>  
 A;Cross-references: GB:U32813; GB:I42023; NID:g1574796; PIDN:AAC22984.1; PID:g1574798; T1  
 A;Note: best homolog was a hypothetical protein from Escherichia coli  
 C;Genetics:  
 A;Note: two copies of this gene are found in the Haemophilus influenzae chromosome  
 C;Superfamily: phosphomannomutase

Query Match 5.3%; Score 7; DB 2; Length 445;  
 Best Local Similarity 100.0%; Pred. No. 53;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 57 ELASAVK 63  
 DB 361 ELASAVK 367

RESULT 18  
 AG1977  
 hypothetical protein alr1370 [imported] - Nostoc sp. (strain PCC 7120)  
 C;Species: Nostoc sp. PCC 7120  
 A;Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120  
 C;Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 09-Dec-2002  
 C;Accession: AG1977  
 R;Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriyuchi,  
 Nakazaki, N.; Shimo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.  
 DNA Res. 8, 205-215, 2001  
 A;Title: Complete Genome Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena  
 A;Reference number: AB1807; MUID:21595285; PMID:11759840  
 A;Accession: AG1977  
 A;Status: preliminary  
 A;Molecule type: DNA  
 A;Residues: 1-454 <NUR>  
 A;Cross-references: GB:BA000019; PIDN:BA073327.1; PID:g17130717; GSPDB:GN00179  
 A;Experimental source: strain PCC 7120  
 C;Genetics:  
 A;Gene: alr1370

Query Match 5.3%; Score 7; DB 2; Length 454;  
 Best Local Similarity 100.0%; Pred. No. 54;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 27 APPREVV 33  
 DB 17 APPREVV 23

RESULT 19  
 D72461

hypothetical protein APE2336 - Aeropyrum pernix (strain K1)  
 C:Species: Aeropyrum pernix  
 C>Date: 20-Aug-1999 #sequence\_revision 20-Aug-1999 #text\_change 20-Aug-1999  
 C/Accession: D72461  
 R:Kawarabayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Halkawa, Y.; Jin-no, K.; Takenawa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; K DNA Res. 6, 83-101, 1999  
 A>Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aeropyrum  
 A:Reference number: A72450, NCID:99310339, PMID:10382966  
 A:Accession: D72461  
 A>Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-467 <RAM>  
 A:Cross-references: DDBJ:AP000064; NID:G5105945; PIDN:BA81348.1; PID:d1045134; PID:g510  
 A:Experimental source: strain K1  
 C:Genetics:  
 A:Gene: APE2336

Query Match 5.3%; Score 7; DB 2; Length 467;  
 Best Local Similarity 100.0%; Pred. No. 55;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 35 PGSGVRI 41  
 |||||  
 Db 157 PGSGVRI 163

RESULT 20  
 T28052  
 hypothetical protein ZK858.7 - Caenorhabditis elegans  
 C:Species: Caenorhabditis elegans  
 C>Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 29-Oct-1999  
 C/Accession: T28052  
 R:White, S.  
 submitted to the EMBL Data Library, September 1996  
 A:Reference number: Z20462  
 A:Accession: T28052  
 A>Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-473 <WIL>  
 A:Cross-references: EMBL:Z79759; PIDN:CAE02135.1; GSPDB:GN00019; CESP:ZK858.7  
 A:Experimental source: clone ZK858  
 C:Genetics:  
 A:Gene: CESP:ZK858.7  
 A:Map position: 1  
 A:Intons: 23/3; 51/2; 177/3; 272/3; 338/3; 404/3; 456/3

Query Match 5.3%; Score 7; DB 2; Length 473;  
 Best Local Similarity 100.0%; Pred. No. 56;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 74 RLGGTGA 80  
 |||||  
 Db 260 RLGGTGA 266

RESULT 21  
 AF3373  
 glutathione-disulfide reductase (EC 1.8.1.7) - Brucella melitensis (strain 16M)  
 C:Species: Brucella melitensis  
 C>Date: 01-Feb-2002 #sequence\_revision 01-Feb-2002 #text\_change 03-May-2002  
 C/Accession: AF3373  
 R:DelVecchio, V.G.; Kaparal, V.; Redkar, R.J.; Patra, G.; Mijer, C.; Los, T.; Ivanova, R.; Mazur, M.; Goldsmid, E.; Selkov, E.; Elzer, F.H.; Hagius, S.; O'Callaghan, D.; Letessier Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002  
 A>Title: The genome sequence of the facultative intracellular pathogen Brucella melitensis  
 A:Reference number: AD3252; PMID:11756688  
 A:Accession: AF3373  
 A>Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-483 <KUR>  
 A:Cross-references: GB:AE008917; PIDN:AAU52153.1; PID:g17982911; GSPDB:GN00190  
 A:Experimental source: strain 16M

C:Genetics:  
 A:Gene: BME10972  
 A:Map position: 1  
 C:Superfamily: dihydrolipoamide dehydrogenase; dihydrolipoamide dehydrogenase homology  
 C:Keywords: oxidoreductase; redox-active disulfide  
 F:62-67/Disulfide Bonds: redox-active #status predicted

Query Match 5.3%; Score 7; DB 2; Length 483;  
 Best Local Similarity 100.0%; Pred. No. 57;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 113 SNGETLE 119  
 |||||  
 Db 269 SNGETLE 275

RESULT 22  
 T36098  
 probable integral membrane protein - Streptomyces coelicolor  
 C:Species: Streptomyces coelicolor  
 C>Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 03-Dec-1999  
 C/Accession: T36098  
 R:Saunders, D.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrett, B.G.; Rajandream, M.A. submitted to the EMBL Data Library, April 1999  
 A:Reference number: Z21596  
 A:Accession: T36098  
 A>Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-488 <SAU>  
 A:Cross-references: EMBL:AL049661; PIDN:CA41214.1; GSPDB:GN00070; SCOEDB:SC134.16  
 A:Experimental source: strain A3(2)  
 C:Genetics:  
 A:Gene: SCOEDB:SC134.16

Query Match 5.3%; Score 7; DB 2; Length 488;  
 Best Local Similarity 100.0%; Pred. No. 57;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 10 PEAEPAA 16  
 |||||  
 Db 313 PEAEPAA 319

RESULT 23  
 B84266  
 cysteinyl-tRNA synthetase [imported] - Halobacterium sp. NRC-1  
 C:Species: Halobacterium sp. NRC-1  
 C>Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 16-Feb-2001  
 C/Accession: B84266  
 R:Ng, W.V.; Kennedy, S.P.; Mahafas, G.G.; Bergquist, B.; Pan, M.; Shukla, H.D.; Lasky, S.; Leitman, E.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Jabor Jung, K.H.; Alam, M.; Freilich, T.  
 Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000  
 A:Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ehardt, H.; Lowe, T.M.; Lie A>Title: Genome sequence of Halobacterium species NRC-1.  
 A:Reference number: AB4160; NCID:20504483; PMID:11016950  
 A:Accession: B84266  
 A>Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-494 <STO>  
 A:Cross-references: GB:AE004437; NID:G10580644; PIDN:AAI9494.1; GSPDB:GN00138  
 C:Genetics:  
 A:Gene: cyes  
 C:Superfamily: cysteine-tRNA ligase

Query Match 5.3%; Score 7; DB 2; Length 494;  
 Best Local Similarity 100.0%; Pred. No. 58;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 56 LELASAV 62  
 |||||  
 Db 394 LELASAV 400



RESULT 24  
B90446  
ABC transporter, permease SSO2714 [imported] - *Sulfolobus solfataricus*  
C/Species: *Sulfolobus solfataricus*  
C/Date: 24-May-2001 #sequence\_revision 24-May-2001 #text\_change 24-May-2001  
C/Accession: B90446  
R/Shu, Q.; Singh, R.K.; Confalonieri, F.; Zivanovic, Y.; Allard, G.; Aways, M.J.; Chan-  
Jong, I.; Jeffries, A.C.; Kozera, C.J.; Medina, N.; Peng, X.; Thi-Ngoc, H.P.; Redder, R.  
arrett, R.A.; Ragan, M.A.; Sersen, C.W.; Van der Oost, J.  
submitted to Genbank, April 2001  
A/Description: *Sulfolobus solfataricus* complete genome.  
A/Reference number: A99139  
A/Accession: B90446  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-548 <NR>  
A/Cross-references: GB:AE006641; NID:G13816034; PIDN:AAK42825.1; GSPDB:GN00155  
C/Genetic:  
A/Map position: 14  
A/Gene: SSO2714

Query Match  
Best Local Similarity 100.0%; Pred. No. 63;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 25 SVAPPE 31  
DB 40 SVAPPE 46

RESULT 25  
T48453  
hypothetical protein T32M21.170 - *Arabidopsis thaliana*  
C/Species: *Arabidopsis thaliana* (mouse-ear cress)  
C/Date: 20-Apr-2000 #sequence\_revision 20-Apr-2000 #text\_change 20-Apr-2000  
C/Accession: T48453  
R/Bevan, M.; Terry, N.; Ardiles, W.; Buyschaert, C.; Daeseville, R.; De Clerck, R.; De  
ewes, H.W.; Rudd, S.; Lemcke, K.; Mayer, K.F.X.  
submitted to the Protein Sequence Database, March 2000  
A/Reference number: Z24487  
A/Accession: T48453  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-555 <BEV>  
A/Cross-references: EMBL:AL162875  
A/Experimental source: cultivar Columbia; BAC clone T32M21  
C/Genetic:  
A/Map position: 5  
A/Introns: 469/3; 496/2; 524/3  
A/Note: T32M21.170

Query Match  
Best Local Similarity 100.0%; Pred. No. 64;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 107 EAIRRAS 113  
DB 441 EAIRRAS 447

RESULT 26  
T37738  
hypothetical protein SPAC167.05 - fission yeast (*Schizosaccharomyces pombe*)  
C/Species: *Schizosaccharomyces pombe*  
C/Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 20-Oct-2000  
C/Accession: T37738; T38953  
R/Rieger, M.; Wood, V.; Rajandream, M.A.; Barrell, B.G.  
submitted to the EMBL Data Library, January 1999  
A/Reference number: Z21743  
A/Accession: T37738  
A/Molecule type: DNA  
A/Residues: 1-601 <RIB>  
A/Cross-references: EMBL:AL035248; PIDN:CAA22850.1; GSPDB:GN00066; SPDB:SPAC167.05

A/Experimental source: strain 972h; cosmid c167  
R/Skelton, J.; Churcher, C.M.; Barrell, B.G.; Rajandream, M.A.; Wood, V.  
submitted to the EMBL Data Library, May 1997  
A/Reference number: Z21819  
A/Accession: T38953  
A/Molecule type: DNA  
A/Residues: 471-601 <SKE>  
A/Cross-references: EMBL:Z95396; PIDN:CAB08759.1; GSPDB:GN00066; SPDB:SPAC57A.01  
C/Experimental source: strain 972h; cosmid c57A7  
C/Genetic:  
A/Map position: 1  
A/Gene: SPDB:SPAC57A.01; SPDB:SPAC167.05

Query Match  
Best Local Similarity 100.0%; Pred. No. 69;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 116 ETLKKT 122  
DB 487 ETLKKT 493

RESULT 27  
A70669  
probable acid-CoA ligase (EC 6.2.1.-) fadD22 [similarity] - *Mycobacterium tuberculosis* (H37Rv)  
N/Alternate names: hypothetical protein pX002a  
C/Species: *Mycobacterium tuberculosis*  
C/Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 03-Nov-2000  
C/Accession: A70669; S73073  
R/Cole, S.T.; Broesch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.;  
Connor, R.; Davies, R.; Devlin, K.; Fellwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.;  
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.  
Nature 393, 537-544, 1998  
A/Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
A/Title: Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome  
A/Reference number: A70500; MUID:98295987; PMID:9634230  
A/Accession: A70669  
A/Status: nucleic acid sequence not shown; translation not shown  
A/Molecule type: DNA  
A/Residues: 1-705 <COL>  
A/Cross-references: GB:Z83858; GB:AL123456; NID:G3261675; PIDN:CAB06101.1; PID:E301287; I  
A/Experimental source: strain H37Rv  
R/Smith, D.R.; Robison, K.  
submitted to the EMBL Data Library, September 1994  
A/Description: *Mycobacterium tuberculosis* cosmid tb02.  
A/Reference number: S73053  
A/Accession: S73073  
A/Molecule type: DNA  
A/Residues: 1-486; 'OO', 489-705 <SMI>  
A/Cross-references: EMBL:U00024; NID:G560506; PIDN:AAA50930.1; PID:G560509  
C/Genetic:  
A/Gene: fadD22  
C/Suprafamily: acyl carrier protein homology; acetate-CoA ligase homology  
C/Keywords: acid-thiol ligase; carrier protein; phosphopantetheine; phosphoprotein  
F:51-486/Domain: acetate-CoA ligase homology <ACI>  
F:544-615/Domain: acyl carrier protein homology <ACPL>  
F:579/Binding site: phosphopantetheine (Ser) (covalent) #status predicted

Query Match  
Best Local Similarity 100.0%; Pred. No. 79;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 34 EFGSGVR 40  
DB 502 EFGSGVR 508

RESULT 28  
CS3292  
penicillin-binding protein 2B (cell-division septum) pbpb - *Bacillus subtilis*  
C/Species: *Bacillus subtilis*  
C/Date: 13-Sep-1995 #sequence\_revision 13-Sep-1995 #text\_change 15-Oct-1999  
C/Accession: CS3292; S43862; F69672

R.Yanouri, A.; Daniel, R.A.; Errington, J.; Buchanan, C.E.  
 J. Bacteriol. 175, 7604-7616, 1993  
 A>Title: Cloning and sequencing of the cell division gene pbbp, which encodes penicillin  
 A/Reference number: A53292; MUID:94064553; PMID:8244529  
 A/Accession: C53292  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-716 <YAN>  
 A/Cross-references: GB:I09703; NID:G304162; PIDN:AA06837.1; PID:G304165  
 R.Daniel, R.A.; Drake, S.; Buchanan, C.E.; Scholte, R.; Errington, J.  
 J. Mol. Biol. 235, 209-220, 1994  
 A>Title: The *Bacillus subtilis* spoVD gene encodes a mother-cell-specific penicillin-band  
 A/Reference number: S43862; MUID:94118264; PMID:8289242  
 A/Accession: S43862  
 A/Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A/Molecule type: DNA  
 A/Residues: 677-716 <DAN>  
 A/Cross-references: EMBL:Z25865; NID:G397893; PIDN:CAA1084.1; PID:G397894  
 A/Note: The nucleotide sequence was submitted to the EMBL Data Library, September 1993  
 R.Kunet, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berth  
 C.; Bron, S.; Brouillet, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Chd  
 A.; Enllich, S.D.; Emmerston, P.T.; Entian, K.D.; Errington, J.; Fader, C.; Ferrari, E.  
 Nature 390, 249-256, 1997  
 A/Authors: Foulger, D.; Fritze, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gall  
 Iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holappel, S.; Hosono, S.; Hullo, M.P.  
 Koester, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Landin  
 A/Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maue  
 Y, M.; Ogawa, K.; Ogilwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portelle  
 Rieger, M.; Rivolta, C.; Rocha, B.; Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Scanlon  
 A/Authors: Scholte, S.; Schroeter, R.; Scottone, F.; Sekiguchi, J.; Sekowska, A.; Seron  
 akuch, M.; Tamakoshi, A.; Tanaka, T.; Terpestra, P.; Tononi, A.; Toesato, V.; Uchiyama,  
 T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K  
 A/Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.  
 A/Title: The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*.  
 A/Reference number: A69580; MUID:98044033; PMID:9384377  
 A/Accession: F69672  
 A/Status: nucleic acid sequence not shown; translation not shown  
 A/Molecule type: DNA  
 A/Residues: 1-716 <KUN>  
 A/Cross-references: GB:299111; GB:AL009126; NID:G2633699; PIDN:CA01389.1; PID:el185106;  
 A/Experimental source: strain 168  
 C/Genetics:  
 A/Genes: pbbp  
 C/Keywords: membrane protein

Query Match 5.3%; Score 7; DB 2; Length 716;  
 Best Local Similarity 100.0%; Pred. No. 80;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 61 AVKEQYP 67  
 Db 633 AVKEQYP 639

RESULT 29  
 D71471  
 Probable DNA mismatch repair - *Chlamydia trachomatis* (serotype D, strain UM3/Cx)  
 C/Species: *Chlamydia trachomatis*  
 C/Date: 13-Sep-1998 #sequence\_revision 13-Sep-1998 #text\_change 08-Oct-1999  
 C/Accession: D71471  
 R/Stephens, R.S.; Kalman, S.; Lammell, C.J.; Fan, J.; Marathe, R.; Aravind, L.; Mitchell,  
 Science 282, 754-759, 1998  
 A/Title: Genome sequence of an obligate intracellular pathogen of humans: *Chlamydia trac*  
 A/Reference number: A71570; MUID:9900809; PMID:9784136  
 A/Accession: D71471  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-820 <ARN>  
 A/Cross-references: GB:AE001351; GB:AE001273; NID:G332247; PIDN:AA06837.1; PID:G332925  
 A/Experimental source: serotype D, strain UM-3/Cx  
 C/Genetics:  
 A/Genes: muts  
 C/Superfamily: DNA mismatch repair protein muts

Query Match 5.3%; Score 7; DB 2; Length 820;  
 Best Local Similarity 100.0%; Pred. No. 90;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 103 KDLIEAI 109  
 Db 154 KDLIEAI 160

RESULT 30  
 A38523  
 genome polypeptide - louping ill virus (strain 369/72) (fragment)  
 N/Contains: capsid protein C; envelope protein E; membrane protein M; nonstructural prote  
 C/Species: louping ill virus  
 C/Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 17-Nov-2000  
 C/Accession: A38523  
 R/Shiu, S.Y.W.; Ayres, M.D.; Gould, E.A.  
 Virology 180, 411-415, 1991  
 A/Title: Genomic sequence of the structural proteins of louping ill virus: comparative ar  
 A/Reference number: A38523; MUID:91082437; PMID:1845834  
 A/Accession: A38523  
 A/Molecule type: genomic RNA  
 A/Residues: 1-960 <SHI>  
 A/Cross-references: GB:M59376; GB:M37687; NID:G331406; PIDN:AAA46281.1; PID:G331407  
 C/Superfamily: Yellow fever virus genome polypeptide  
 C/Keywords: capsid protein; envelope protein; glycoprotein; nonstructural protein; poly  
 F/2-116/Product: capsid protein C #status predicted <CPC>  
 F/101-119/Domain: transmembrane #status predicted <TM1>  
 F/117-280/Product: membrane protein M precursor #status predicted <MP>  
 F/117-280/Domain: nonterminal signal sequence #status predicted <SIG>  
 F/206-280/Product: membrane protein M #status predicted <MP>  
 F/246-264/Domain: transmembrane #status predicted <TM2>  
 F/266-286/Domain: transmembrane #status predicted <TM3>  
 F/281-776/Product: envelope protein E #status predicted <EP>  
 F/716-736/Domain: transmembrane #status predicted <TM4>  
 F/760-776/Domain: transmembrane #status predicted <TM5>  
 F/777-980/Product: nonstructural protein NS1 (fragment) #status predicted <NS1>  
 F/144,434,641,753,861/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 5.3%; Score 7; DB 2; Length 960;  
 Best Local Similarity 100.0%; Pred. No. 11e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 58 LASAYKE 64  
 Db 818 LASAYKE 824

RESULT 31  
 T19048  
 Probable Pro-X carboxypeptidase F23B2.12 - *Caenorhabditis elegans*  
 C/Species: *Caenorhabditis elegans*  
 C/Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 29-Oct-1999  
 C/Accession: T19048; T21307  
 R/McMurray, A.  
 submitted to the EMBL Data Library, December 1995  
 A/Reference number: Z19065  
 A/Accession: T19048  
 A/Status: preliminary; translated from GB/EMBL/DBJ  
 A/Molecule type: DNA  
 A/Residues: 1-1080 <MIL>  
 A/Cross-references: EMBL:Z68295; PIDN:CAA92588.1; GSPDB:GN00022; CESP:F23B2.12  
 A/Experimental source: clone C07C7  
 R/McMurray, A.  
 submitted to the EMBL Data Library, November 1996  
 A/Reference number: Z19403  
 A/Accession: T21307  
 A/Status: preliminary; translated from GB/EMBL/DBJ  
 A/Molecule type: DNA  
 A/Residues: 1-1080 <M12>  
 A/Cross-references: EMBL:Z82266; PIDN:CA05187.1; GSPDB:GN00022; CESP:F23B2.12  
 A/Experimental source: clone F23B2

C:Genetics:  
 A:Gene: CESP.F23B2.12  
 A:Map position: 4  
 A:Introns: 19/2; 127/2; 287/3; 349/1; 658/2; 916/2  
 C:Keywords: duplication

Query Match  
 Best Local Similarity 5.3%; Score 7; DB 2; Length 1080;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Db 527 FSKLENG 533

RESULT 32  
 A57650  
 repair protein XRG - mouse  
 C:Species: Mus musculus (house mouse)  
 C:Date: 08-Feb-1996 #sequence\_revision 08-Feb-1996 #text\_change 05-Nov-1999  
 C:Accession: A57650  
 R:Harada, Y.N.; Matsuda, Y.; Shiomi, N.; Shiomi, T.  
 Genomics 28, 59-65, 1995  
 A:Title: Complementary DNA sequence and chromosomal localization of xpg, the mouse count  
 A:Reference number: A57650; MUID:96070433; PMID:7590748  
 A:Accession: A57650  
 A:Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-1170 <HAR>  
 A:Cross-references: GB:D16306; NID:G1389576; PIDN:BA003813.1; PID:d1004329; PID:g1389577  
 C:Keywords: DNA repair

Query Match  
 Best Local Similarity 5.3%; Score 7; DB 2; Length 1170;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Db 58 LASAVKE 64  
 486 LASAVKE 492

RESULT 33  
 P00667  
 photosystem I 19.3K D1 chain - common tobacco (fragment)  
 C:Species: Nicotiana tabacum (common tobacco)  
 C:Date: 19-May-1994 #sequence\_revision 19-May-1994 #text\_change 17-Mar-1999  
 C:Accession: P00667  
 R:Obokata, J.; Mikami, K.; Hayashida, N.; Nakamura, M.; Sugitara, M.  
 Plant Physiol. 102, 1259-1267, 1993  
 A:Title: Molecular heterogeneity of photosystem I. psad, psaf, psag and psal are a  
 A:Reference number: P00667; MUID:94105345; PMID:8278548  
 A:Accession: P00667  
 A:Molecule type: protein  
 A:Residues: 1-22 <OBO>  
 C:Keywords: chloroplast; photosynthesis; photosystem I; thylakoid

Query Match  
 Best Local Similarity 4.6%; Score 6; DB 2; Length 22;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Db 11 EAPPA 16  
 11 EAPPA 16

RESULT 34  
 B97906  
 phospho-beta-glucosidase, truncation [imported] - Streptococcus pneumoniae (strain R6)  
 C:Species: Streptococcus pneumoniae  
 C:Date: 22-Oct-2001 #sequence\_revision 22-Oct-2001 #text\_change 22-Oct-2001  
 C:Accession: B97906

R:Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Burgett, S.; DeHoff, B.S.; Es  
 e, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.; Mc  
 Y, P.; Sun, P.M.; Winkler, M.E.  
 J. Bacteriol. 183, 5709-5717, 2001  
 A:Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R.;  
 A:Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.  
 A:Reference number: A97872; MUID:21429245; PMID:11544234  
 A:Accession: B97906

A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-30 <KUR>  
 A:Cross-references: GB:AE007317; PIDN:AAK9078.1; PID:g15457825; GSPDB:GN00174  
 C:Genetics:  
 A:Gene: bg1B-truncation

Query Match  
 Best Local Similarity 4.6%; Score 6; DB 2; Length 30;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Db 113 SNGERTL 118  
 23 SNGERTL 28

RESULT 35  
 D69466  
 conserved hypothetical protein AF1733 - Archaeoglobus fulgidus  
 C:Species: Archaeoglobus fulgidus  
 C:Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 21-Jul-2000  
 C:Accession: D69466  
 R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson,  
 J.; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F.;  
 Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.  
 Nature 390, 364-370, 1997  
 A:Authors: Utterback, T.; Cotton, M.D.; Spriggs, T.; Attiach, P.; Kaine, B.P.; Sykes, S.N.  
 Smith, H.O.; Weese, C.R.; Venter, J.C.  
 A:Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archaeo  
 A:Reference number: A69250; MUID:98049343; PMID:9389475  
 A:Accession: D69466  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-45 <KLB>  
 A:Cross-references: GB:AE000983; GB:AE00782; NID:g2689306; PIDN:AA89518.1; PID:g2648820  
 C:Superfamily: Methanococcus jannaschii conserved hypothetical protein MJ151

Query Match  
 Best Local Similarity 4.6%; Score 6; DB 1; Length 45;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Db 39 VRIIVE 44  
 5 VRIIVE 10

RESULT 36  
 A40424  
 phospholamban - human  
 C:Species: Homo sapiens (man)  
 C:Date: 28-Feb-1992 #sequence\_revision 27-Jun-1994 #text\_change 18-Jun-1999  
 C:Accession: A40424  
 R:Fujii, J.; Zairan-Herzberg, A.; Willard, H.F.; Tada, M.; MacLennan, D.H.  
 J. Biol. Chem. 266, 11669-11675, 1991  
 A:Title: Structure of the rabbit phospholamban gene, cloning of the human cDNA, and assig  
 A:Reference number: A40424; MUID:91286032; PMID:1828805  
 A:Accession: A40424  
 A:Molecule type: mRNA  
 A:Residues: 1-52 <FTU>  
 A:Cross-references: GB:M63603; NID:g189942; PIDN:AAA60083.1; PID:g189943  
 C:Comment: Phospholamban is expressed in cardiac muscle, slow twitch skeletal muscle, and  
 e, after phosphorylation, the Ca++ pump is activated and the rate of muscle relaxation if  
 C:Comment: Helical transmembrane domains of five chains are thought to aggregate in the  
 C:Genetics:

A:Gene: GDB:PLN; PUB  
A:Cross-references: GDB:128300; OMIM:172405  
A:Map position: 6q22.1-6q22.1  
C:Superfamily: phospholamban  
C:Keywords: acetylated amino end; ATPase inhibitor; muscle; pentamer; phosphoprotein; tr  
F/31-52/Domain: transmembrane #status predicted <TM>  
F/1/Modified site: acetylated amino end (Met) #status predicted  
F/16/Binding site: phosphate (Thr) (covalent) (by calmodulin-dependent kinase) #status p  
F/17/Binding site: phosphate (Thr) (covalent) (by calmodulin-dependent kinase) #status p

Query Match 4.6%; Score 6; DB 1; Length 52;  
Best Local Similarity 100.0%; Pred. No. 82;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 108 AIRRAS 113  
|||||  
DB 11 AIRRAS 16

RESULT 37  
S37638  
phospholamban - rat  
C:Species: Rattus norvegicus (Norway rat)  
C>Date: 05-Mar-1994 #sequence revision 27-Jun-1994 #text\_change 18-Jun-1999  
C:Accession: S37638; I52270; I64795; I51840  
R:Shanahan, C.M.; Weisberg, P.L.; Metcalfe, J.C.  
Circ. Res. 73, 193-204, 1993  
A:Title: Isolation of gene markers of differentiated and proliferating vascular smooth m  
A:Reference number: S37637; MUID:93284726; PMID:8508530  
A:Accession: S37638  
A:Molecule type: mRNA  
A:Residues: 1-52 <SHA>  
A:Cross-references: EMBL:X71068; NID:G313809; PIDN:CAA50394.1; PID:G313810  
R:Johns, D.C.; Feldman, A.M.  
Biochem. Biophys. Res. Commun. 188, 927-933, 1992  
A:Title: Identification of a highly conserved region at the 5' flank of the phospholamb  
A:Reference number: I52270; MUID:93075183; PMID:1445334  
A:Accession: I52270  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-52 <JOH>  
A:Cross-references: GB:I03382; NID:G206134; PIDN:AAA41849.1; PID:G206136  
R:Wang, K.S.; Nadal-Ginard, B.  
Adv. Exp. Med. Biol. 304, 387-395, 1991  
A:Title: Cloning phospholamban cDNA from rat aortic smooth muscle.  
A:Reference number: I51840; MUID:92206263; PMID:1725098  
A:Accession: I64795  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Status: not compared with conceptual translation  
A:Molecule type: mRNA  
A:Residues: 1-52 <HWA1>  
A:Cross-references: GB:S95853; NID:G247932; PIDN:AAB21903.1; PID:G247933  
A:Accession: I51840  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-52 <HWA2>  
A:Cross-references: GB:S95849; NID:G247934  
C:Comment: Phospholamban is expressed in cardiac muscle, slow twitch skeletal muscle, an  
C:Comment: Phospholamban is the major phosphorylated protein in cardiac muscle sarcoplas  
Pase; after phosphorylation, the calcium pump is activated and the rate of muscle relax  
C:Superfamily: phospholamban  
C:Keywords: acetylated amino end; ATPase inhibitor; cardiac muscle; heart; pentamer; pnc  
F/31-52/Domain: transmembrane #status predicted <TM>  
F/1/Modified site: acetylated amino end (Met) #status predicted  
F/16/Binding site: phosphate (Ser) (covalent) (by CAMP-dependent kinase) #status predict  
F/17/Binding site: phosphate (Thr) (covalent) (by calmodulin-dependent kinase) #status p

RESULT 38  
A49057  
phospholamban - mouse  
C:Species: Mus musculus (house mouse)  
C>Date: 19-Dec-1993 #sequence revision 27-Jun-1994 #text\_change 18-Jun-1999  
C:Accession: A49057  
R:Gantim, J.R.; Luo, W.; Pomiah, S.; Grupp, I.; Kim, H.W.; Ferguson, D.G.; Kadambi, V.; F  
Circ. Res. 71, 1021-1030, 1992  
A:Title: Mouse phospholamban gene expression during development in vivo and in vitro.  
A:Reference number: A49057; MUID:93008802; PMID:1394867  
A:Accession: A49057  
A:Molecule type: mRNA  
A:Residues: 1-52 <GAN>  
A:Cross-references: GB:S46792; NID:G257745; PIDN:AAB23706.1; PID:G257746  
A:Experimental source: cardiac muscle  
A:Note: sequence extracted from NCBI backbone (NCBIN:116999, NCBI:P:117001)  
C:Comment: Phospholamban is expressed in cardiac muscle, slow twitch skeletal muscle, and  
C:Comment: Phospholamban is the major phosphorylated protein in cardiac muscle sarcoplas  
e; after phosphorylation, the Ca++ pump is activated and the rate of muscle relaxation is  
C:Superfamily: phospholamban  
C:Keywords: acetylated amino end; ATPase inhibitor; muscle; pentamer; phosphoprotein; tr  
F/31-52/Domain: transmembrane #status predicted <TM>  
F/1/Modified site: acetylated amino end (Met) #status predicted  
F/16/Binding site: phosphate (Ser) (covalent) (by CAMP-dependent kinase) #status predict  
F/17/Binding site: phosphate (Thr) (covalent) (by calmodulin-dependent kinase) #status p

Query Match 4.6%; Score 6; DB 1; Length 52;  
Best Local Similarity 100.0%; Pred. No. 82;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 108 AIRRAS 113  
|||||  
DB 11 AIRRAS 16

RESULT 39  
B40424  
phospholamban - rabbit  
C:Species: Oryctolagus cuniculus (domestic rabbit)  
C>Date: 28-Feb-1992 #sequence revision 27-Jun-1994 #text\_change 18-Jun-1999  
C:Accession: B40424; S00249  
R:Fujii, J.; Zarain-Herzberg, A.; Willard, H.F.; Tada, M.; MacLennan, D.H.  
U. Biol. Chem. 266, 11659-11675, 1991  
A:Title: Structure of the rabbit phospholamban gene, cloning of the human cDNA, and assa  
A:Reference number: A40424; MUID:91268032; PMID:1828605  
A:Accession: B40424  
A:Molecule type: DNA  
A:Residues: 1-52 <FUJ>  
A:Cross-references: GB:M63601; NID:G165636; PIDN:AAA11445.1; PID:G165639  
R:Fujii, J.; Lytton, J.; Tada, M.; MacLennan, D.H.  
FEBS Lett. 227, 51-55, 1988  
A:Title: Rabbit cardiac and slow-twitch muscle express the same phospholamban gene.  
A:Reference number: S00249; MUID:88112222; PMID:2962883  
A:Accession: S00249  
A:Status: not compared with conceptual translation  
A:Molecule type: mRNA  
A:Residues: 1-52 <FUJ>  
A:Cross-references: GB:Y00761; NID:G1661; PIDN:CAA68730.1; PID:G1662  
C:Comment: Phospholamban is expressed in cardiac muscle, slow twitch skeletal muscle, and  
C:Comment: Phospholamban is the major phosphorylated protein in cardiac muscle sarcoplas  
e; after phosphorylation, the Ca++ pump is activated and the rate of muscle relaxation is  
C:Genetics:  
A:Note: only one gene was detected  
A:Note: the single intron is upstream of the coding region  
C:Superfamily: phospholamban  
C:Keywords: acetylated amino end; ATPase inhibitor; muscle; pentamer; phosphoprotein; tr  
F/31-52/Domain: transmembrane #status predicted <TM>  
F/1/Modified site: acetylated amino end (Met) #status predicted  
F/16/Binding site: phosphate (Ser) (covalent) (by CAMP-dependent kinase) #status predict  
F/17/Binding site: phosphate (Thr) (covalent) (by calmodulin-dependent kinase) #status p

Query Match 4.6%; Score 6; DB 1; Length 52;  
 Best Local Similarity 100.0%; Pred. No. 82;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 108 AIRRAS 113  
 |||||  
 Db 11 AIRRAS 16

# RESULT 40

A29002  
 phospholamban - dog

C/Species: Canis lupus familiaris (dog)

C/Date: 23-Aug-1987 #sequence\_revision 27-Jun-1994 #text\_change 18-Jun-1999

C/Accession: A29002; A26805; A25307; A24818; I46227

R/Fujii, J.; Ueno, A.; Kitano, K.; Tanaka, S.; Kadoma, M.; Tada, M.

J. Clin. Invest. 79, 301-304, 1987

A/Title: Complete complementary DNA-derived amino acid sequence of canine cardiac phospholamban

A/Reference number: A29002; MUID:87083954; PMID:3793929

A/Accession: A29002

A/Molecule type: mRNA

A/Residues: 1-52 <FNU>

A/Cross-references: GB:M16012; NID:g164043; PIDN:AAA30884.1; PID:g164044

R/Ueda, A.; Kitano, K.; Fujii, J.; Kadoma, M.; Tada, M.; Tanaka, S.

Nucleic Acids Res. 15, 6738, 1987

A/Title: The cDNA sequence of the major phospholamban mRNA in canine cardiac ventricular

A/Reference number: A26805; MUID:87316936; PMID:3628007

A/Accession: A26805

A/Molecule type: mRNA

A/Residues: 1-52 <UYE>

A/Cross-references: GB:Y00399; NID:g911; PIDN:CAA66461.1; PID:g912

R/Stimmerman, H.K.B.; Collins, J.H.; Theibert, J.L.; Wegener, A.D.; Jones, L.R.

J. Biol. Chem. 261, 13333-13341, 1986

A/Title: Sequence analysis of phospholamban. Identification of phosphorylation sites and

A/Reference number: A25307; MUID:87008549; PMID:3759968

A/Accession: A25307

A/Contents: partial sequence and phosphorylation sites

A/Accession: A25307

A/Molecule type: protein

A/Residues: 10-45 <SIM>

R/Fujii, J.; Kadoma, M.; Tada, M.; Sakiyama, F.

Biochem. Biophys. Res. Commun. 138, 1044-1050, 1986

A/Title: Characterization of structural unit of phospholamban by amino acid sequencing

A/Reference number: A24818; MUID:8633152; PMID:3753485

A/Accession: A24818

A/Molecule type: protein

A/Residues: 1-35, 'X', '37-40', 'X', '42-45' <FNU>

R/Ueda, A.; Kitano, K.; Fujii, J.; Kadoma, M.; Tada, M.; Tanaka, S.

Nucleic Acids Symp. Ser. 17, 121-124, 1986

A/Title: Characterization of recombinant cDNA clones for canine cardiac phospholamban.

A/Reference number: I46227; MUID:87174860; PMID:3562256

A/Accession: I46227

A/Status: translated from GB/EMBL/DBJ

A/Molecule type: mRNA

A/Residues: 1-52 <UYE>

A/Cross-references: GB:M35393; NID:g164045; PIDN:AA41618.1; PID:g164045

C/Comment: Phospholamban is expressed in cardiac muscle, slow twitch skeletal muscle, and

e; after phosphorylation, the Ca++ pump is activated and the rate of muscle relaxation is

C/Superfamily: phospholamban

C/Keywords: acetylated amino end; ATPase inhibitor; muscle; pentamer; phosphoprotein; tr

F/31-52/Domain: transmembrane #status predicted <TMM>

F/1/Modified site: acetylated amino end (Met) #status experimental

F/16/Binding site: phosphate (Thr) (covalent) (by calmodulin-dependent kinase) #status ex

F/17/Binding site: phosphate (Thr) (covalent) (by calmodulin-dependent kinase) #status e

## Query Match

Best Local Similarity 100.0%; Score 6; DB 1; Length 52;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 108 AIRRAS 113  
 |||||  
 Db 11 AIRRAS 16

# RESULT 41

S05540

phospholamban - pig

C/Species: Sus scrofa domestica (domestic pig)

C/Date: 21-Nov-1993 #sequence\_revision 27-Jun-1994 #text\_change 18-Jun-1999

C/Accession: S05540

R/Verboomen, H.; Muytack, F.; Eggermont, J.A.; de Jaegere, S.; Misaen, L.; Raeymaekers,

Biochem. J. 262, 353-356, 1989

A/Title: cDNA cloning and sequencing of phospholamban from pig stomach smooth muscle.

A/Reference number: S05540; MUID:90056437; PMID:2530978

A/Accession: S05540

A/Molecule type: mRNA

A/Residues: 1-52 <VER>

A/Cross-references: EMBL:X15075; NID:g2055; PIDN:CAA33171.1; PID:g2056

C/Comment: Phospholamban is expressed in cardiac muscle, slow twitch skeletal muscle, and

e; after phosphorylation, the Ca++ pump is activated and the rate of muscle relaxation is

C/Superfamily: phospholamban

C/Keywords: acetylated amino end; ATPase inhibitor; muscle; pentamer; phosphoprotein; tr

F/31-52/Domain: transmembrane #status predicted <TMM>

F/1/Modified site: acetylated amino end (Met) #status predicted

F/16/Binding site: phosphate (Thr) (covalent) (by calmodulin-dependent kinase) #status p

F/17/Binding site: phosphate (Thr) (covalent) (by calmodulin-dependent kinase) #status p

## Query Match

Best Local Similarity 100.0%; Score 6; DB 1; Length 52;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 108 AIRRAS 113  
 |||||  
 Db 11 AIRRAS 16

# RESULT 42

AF2153

hypothetical protein aer2781 [imported] - Nostoc sp. (strain PCC 7120)

C/Species: Nostoc sp. PCC 7120

A/Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120

C/Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 09-Dec-2002

C/Accession: AF2153

R/Kanevo, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriyuchi,

Nakazaki, N.; Shimo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Tabata, S.

DNA Res. 8, 205-213, 2001

A/Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anak

A/Reference number: AB1807; MUID:21595285; PMID:11759840

A/Accession: AF2153

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-56 <KUR>

A/Cross-references: GB:BA000019; PIDN:BA874480.1; PID:g17131874; GSPDB:GN00179

A/Experimental source: strain PCC 7120

C/Genetics:

A/Genes: aer2781

## Query Match

Best Local Similarity 100.0%; Score 6; DB 2; Length 56;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 GFEAP 14  
 |||||  
 Db 24 GFEAP 29

# RESULT 43

AH2604

conserved hypothetical protein Atu0232 [imported] - Agrobacterium tumefaciens (strain C5)

C/Species: Agrobacterium tumefaciens

C/Date: 11-Jan-2002 #sequence\_revision 11-Jan-2002 #text\_change 18-Nov-2002

C/Accession: AH2604

R/Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, L.

erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kuyavlin, T.; Levy, R.; Li, M.; McEllele

; Karp, P.; Romero, P.; Zhang, S.  
 Science 294, 2317-2323, 2001  
 A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,  
 ster, E.W.  
 A>Title: The Genome of the Natural Genetic Engineer *Agrobacterium tumefaciens* C58.  
 A:Reference number: AB2577; MUID:21608550; PMID:11743193  
 A:Accession: AH2604  
 A>Status: Preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-62 <KUR>  
 A:Cross-references: GB:AE008688; PIDN:AA141254.1; PID:g17738560; GSPDB:GN00186  
 A:Experimental source: strain C58 (Dupont)  
 C:Genetics:  
 A:Gene: Atcu0232  
 A:Map position: circular chromosome

Query Match 4.6%; Score 6; DB 2; Length 62;  
 Best Local Similarity 100.0%; Pred. No. 96;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 112 ASNGET 117  
 Db 20 ASNGET 25

RESULT 44  
 G97386  
 hypothetical 12.7K protein in ptx operon 5' region (orf117) [imported] - *Agrobacterium t*  
 C/Species: *Agrobacterium tumefaciens*  
 C/Date: 30-Sep-2001 #sequence\_revision 30-Sep-2001 #text\_change 18-Nov-2002  
 C/Accession: G97386  
 R/Gooder, B.; Hinkle, G.; Gatung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman,  
 A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.;  
 Science 294, 2323-2328, 2001  
 A>Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent *Agrobacterium tum*  
 A:Reference number: A97359; MUID:21608551; PMID:11743194  
 A:Accession: G97386  
 A:Molecule type: DNA  
 A:Residues: 1-62 <KUR>  
 A:Cross-references: GB:AE007869; PIDN:AAK86048.1; PID:g15155121; GSPDB:GN00169  
 C:Genetics:  
 A:Gene: AGR\_C\_390  
 A:Map position: circular chromosome

Query Match 4.6%; Score 6; DB 2; Length 62;  
 Best Local Similarity 100.0%; Pred. No. 96;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 112 ASNGET 117  
 Db 20 ASNGET 25

RESULT 45  
 JH0338  
 T-cell receptor alpha chain V region (V2.1) - mouse (fragment)  
 C/Species: *Mus musculus* (house mouse)  
 C/Date: 12-Feb-1993 #sequence\_revision 12-Feb-1993 #text\_change 23-Jul-1999  
 C/Accession: JH0338  
 R/Taylor, A.H.; Haberman, A.M.; Gerhard, W.; Caton, A.J.  
 J. Exp. Med. 172, 1643-1651, 1990  
 A>Title: Structure-function relationships among highly diverse T cells that recognize a  
 A:Reference number: JH0333; MUID:91079767; PMID:1701821  
 A:Accession: JH0338  
 A:Molecule type: mRNA  
 A:Residues: 1-64 <TAY>  
 A:Cross-references: GB:MA4204; NID:g201177; PIDN:AA58752.1; PID:g201178  
 C/Comment: This T-cell receptor recognizes a determinant from influenza virus hemagglut  
 C:Superfamily: immunoglobulin V region; immunoglobulin homology  
 C/Keywords: receptor; T-cell

Query Match 4.6%; Score 6; DB 2; Length 64;

Best Local Similarity 100.0%; Pred. No. 98;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 50 GPEATY 55  
 Db 15 GPEATY 20

RESULT 46  
 AI3203  
 conserved hypothetical protein Atcu5359 [imported] - *Agrobacterium tumefaciens* (strain C5  
 C/Species: *Agrobacterium tumefaciens*  
 C/Date: 11-Jan-2002 #sequence\_revision 11-Jan-2002 #text\_change 18-Nov-2002  
 C/Accession: AI3203  
 R/Wood, D.W.; Serubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, L.  
 erage, G.; Giller, W.; Grant, C.; Guenther, D.; Kutyavln, T.; Levy, R.; Li, M.; McCellie  
 ; Karp, P.; Romero, P.; Zhang, S.  
 Science 294, 2317-2323, 2001  
 A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm, E  
 ster, E.W.  
 A>Title: The Genome of the Natural Genetic Engineer *Agrobacterium tumefaciens* C58.  
 A:Reference number: AB2577; MUID:21608550; PMID:11743193  
 A:Accession: AI3203  
 A>Status: Preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-67 <KUR>  
 A:Cross-references: GB:AE008687; PIDN:AA146047.1; PID:g17743806; GSPDB:GN00188  
 A:Experimental source: strain C58 (Dupont)  
 C:Genetics:  
 A:Gene: Atcu5359  
 A:Genome: plasmid

Query Match 4.6%; Score 6; DB 2; Length 67;  
 Best Local Similarity 100.0%; Pred. No. 1e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 111 RASNGE 116  
 Db 24 RASNGE 29

RESULT 47  
 C42266  
 peptidylglycine monooxygenase (EC 1.14.17.3), rPAM-3b - rat (fragment)  
 N/Alternate names: peptidylglycine alpha-amidating monooxygenase  
 C/Species: *Rattus norvegicus* (Norway rat)  
 C/Date: 04-Mar-1993 #sequence\_revision 18-Nov-1994 #text\_change 26-May-1995  
 C/Accession: C42266  
 R/Eipper, B.A.; Green, C.B.; Campbell, T.A.; Stoffers, D.A.; Keutmann, H.T.; Mains, R.E.;  
 J. Biol. Chem. 267, 4008-4015, 1992  
 A>Title: Alternative splicing and endoproteolytic processing generate tissue-specific for  
 A:Reference number: A42266; MUID:92156145; PMID:1140449  
 A:Accession: C42266  
 A>Status: Preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-74 <EIP>  
 A:Experimental source: pituitary  
 A>Note: sequence extracted from NCBI backbone (NCIN:83167, NCBI:83169)  
 C:Superfamily: peptidylglycine monooxygenase II; peptidylglycine monooxygenase I homolog  
 C/Keywords: oxidoreductase

Query Match 4.6%; Score 6; DB 2; Length 74;  
 Best Local Similarity 100.0%; Pred. No. 1e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 34 EPSSGV 39  
 Db 33 EPSSGV 38

RESULT 48  
 D87654  
 transcription regulator, Cro/ci family [imported] - *Caulobacter crescentus*

C:Species: *Caulobacter crescentus*  
 C>Date: 20-Apr-2001 #sequence\_revision 20-Apr-2001 #text\_change 20-Apr-2001  
 C/Accession: D87654  
 R:Merhan, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.B.; Laub, M.T.; Debboy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Hatt, D.H.; Kolor, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M. Proc. Natl. Acad. Sci. U.S.A. 98, 4138-4141, 2001  
 A>Title: Complete Genome Sequence of *Caulobacter crescentus*.  
 A/Reference number: A87249; MUID:21173698; PMID:11559647  
 A/Accession: D87654  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-75 <STD>  
 A/Cross-references: GB:AE005673; NID:G13424960; PIDN:AAK25232.1; GSPDB:GN00148  
 C/Genetics:  
 A:Gene: CC3270

Query Match 4.6%; Score 6; DB 2; Length 75;  
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 116 ETEKXI 121  
 DB 46 ETEKXI 51

RESULT 49  
 A11634  
 protein encoded by Thp16 homolog l1n1618 [imported] - *Listeria innocua* (strain Clp11262)  
 C/Species: *Listeria innocua*  
 C>Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 27-Nov-2001  
 C/Accession: A11634  
 R:Glaser, P.; Frangoul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker, D.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dusurget, O.; Entian, K.D.; Fsihi, H.; Jones, L.M.; Karst, U. Science 294, 849-852, 2001  
 A/Authors: Kreft, U.; Kuhn, M.; Kunst, F.; Kurapkac, G.; Madueno, E.; Maitournam, A.; Mak, C.; Schlietner, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Weiland, A./Title: Comparative genomics of *Listeria* species.  
 A/Reference number: AB1077; MUID:21537279; PMID:11679669  
 A/Accession: A11634  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-75 <GLA>  
 A/Cross-references: GB:AL592022; PIDN:CAC6849.1; PID:G16414105; GSPDB:GN00178  
 A:Experimental source: strain Clp11262  
 C/Genetics:  
 A:Gene: l1n1618

Query Match 4.6%; Score 6; DB 2; Length 75;  
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 104 DLIEAI 109  
 DB 64 DLIEAI 69

RESULT 50  
 A43537  
 heat-stable antigen M1/69-J11d precursor - mouse  
 N/Alternate names: CD24 protein; necardin  
 C/Species: *Mus musculus* (house mouse)  
 C>Date: 06-Nov-1992 #sequence\_revision 06-Nov-1992 #text\_change 31-Jan-2000  
 C/Accession: A43537; I48287; S15784; S15783; S43709; S32240; S33129  
 R:Kay, R.; Takel, F.; Humphries, R.K. J. Immunol. 145, 1952-1959, 1990  
 A>Title: Expression cloning of a cDNA encoding M1/69-J11d heat-stable antigens.  
 A/Reference number: A43537; MUID:90361906; PMID:2118158  
 A/Accession: A43537  
 A/Molecule type: mRNA  
 A/Residues: 1-76 <KAY>  
 A/Cross-references: GB:M58661; NID:G198985; PIDN:AAA39481.1; PID:G198986

R;Wenger, R.H.; Rochelle, J.M.; Seidin, M.F.; Kohler, G.; Nielsen, P.J. J. Biol. Chem. 268, 23345-23352, 1993  
 A>Title: The heat stable antigen (mouse CD24) gene is differentially regulated but has a

A/Reference number: A48876; MUID:94043127; PMID:8226859  
 A/Accession: I48287  
 A/Status: translation not shown; translated from GB/EMBL/DBJ  
 A/Molecule type: DNA  
 A/Residues: 1-76 <RSG>  
 A/Cross-references: EMBL:X72910; NID:G296466; PIDN:CAA51415.1; PID:G296467  
 R;Wenger, R.H.; Ayane, M.; Bose, R.; Koehler, G.; Nielsen, P.J. Eur. J. Immunol. 21, 1039-1046, 1991

A/Title: The genes for a mouse hematopoietic differentiation marker called the heat-stab)  
 A/Reference number: S15783; MUID:91209380; PMID:2019286  
 A/Accession: S15784  
 A/Molecule type: DNA  
 A/Residues: 1-76 <WE3>

A/Cross-references: EMBL:X56469; NID:G51439; PIDN:CAA39841.1; PID:G51440  
 A/Accession: S15783  
 A/Status: nucleic acid sequence not shown  
 A/Molecule type: mRNA  
 A/Residues: 32-76 <WE2>

A/Cross-references: EMBL:X53825  
 R;Nielsen, P.J. submitted to the EMBL Data Library, July 1990  
 A/Reference number: S19111  
 A/Accession: S43709  
 A/Molecule type: mRNA  
 A/Residues: 1-76 <NIE>

A/Cross-references: EMBL:X53825; NID:G55441; PIDN:CAA37822.1; PID:G55442  
 C/Genetics:  
 A:Gene: Cd24a  
 A/Map position: 10  
 A/Intons: 23/3

C/Keywords: blocked carboxyl end; glycoprotein; lipoprotein; membrane protein; phosphatic  
 F;1-26/Domain: signal sequence #status predicted <SIG>  
 F;27-56/Product: heat-stable antigen M1/69-J11d #status predicted <MAT>  
 F;57-76/Domain: carboxyl-terminal propeptide #status predicted <CTP>  
 F;56/Modified site: GPI-anchor ethanolamine amidated carboxyl end (Ser) (in mature form)

Query Match 4.6%; Score 6; DB 2; Length 76;  
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 23 QTSVAP 28  
 DB 28 QTSVAP 33

Search completed: December 15, 2003, 17:07:21  
 Job time : 24 secs

**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 15, 2003, 17:00:21 ; Search time 17 Seconds  
(without alignments)  
362.382 Million cell updates/sec

Title: US-09-925-301-966

Perfect score: 131  
Sequence: 1 AEVHTRKQGPBAEPAMSGE.....ASNGETLEKITSRPPCVIL 131

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 127863 seqs, 47026705 residues

Word size : 0

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 100 summaries

Database : SwissProt 41.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	9	6.9	158	1	VE6_HPV45
2	8	6.1	220	1	Y132_METUA
3	7	5.3	78	1	DLTC_LISIN
4	7	5.3	88	1	ATP9_DICDI
5	7	5.3	122	1	R18E_PVRAB
6	7	5.3	160	1	P164_MYCTU
7	7	5.3	217	1	ENGB_AGRTS
8	7	5.3	245	1	TEV1_BPT4
9	7	5.3	353	1	PMT2_TOBAC
10	7	5.3	375	1	PMT1_TOBAC
11	7	5.3	377	1	PRLP_RAT
12	7	5.3	378	1	PRLP_MOUSE
13	7	5.3	381	1	PMT3_TOBAC
14	7	5.3	382	1	PRLP_BOVIN
15	7	5.3	382	1	PRLP_HUMAN
16	7	5.3	419	1	PMT4_TOBAC
17	7	5.3	445	1	MESA_HAEIN
18	7	5.3	494	1	SYC_HAINT
19	7	5.3	579	1	COE2_BEAR
20	7	5.3	601	1	YFKS_SCHPO
21	7	5.3	716	1	PBPB_BACSV
22	7	5.3	820	1	MUTS_CHLTR
23	7	5.3	980	1	POIG_LIV
24	7	5.3	1170	1	XPG_MOUSE
25	7	5.3	1729	1	DME_ARATH
26	7	5.3	2353	1	COAH_HUMAN
27	6	4.6	45	1	YH33_ARCFU
28	6	4.6	52	1	PPLA_HUMAN
29	6	4.6	52	1	PPLA_MOUSE
30	6	4.6	52	1	PPLA_PIG
31	6	4.6	56	1	RK33_GUTH
32	6	4.6	76	1	CD24_MOUSE
33	6	4.6	76	1	CD24_RAT

RESULT 1

34	6	4.6	83	1	PYS1_SYNY3	P73202 synechocyst
35	6	4.6	102	1	YAB5_YEAST	P39725 saccharomyc
36	6	4.6	112	1	RS6_CHLMU	Q91c1l chlamydia m
37	6	4.6	112	1	RS6_CHLTR	Q84807 chlamydia t
38	6	4.6	115	1	TAT_STVGB	P22384 simian immu
39	6	4.6	116	1	TAT_STVAM	P36340 simian immu
40	6	4.6	120	1	R18B_PYRFU	Q80e5 pyrococcus
41	6	4.6	120	1	RL18_BACHD	Q80e5 pyrococcus
42	6	4.6	126	1	CYCP_RHURU	P00144 rhodospiril
43	6	4.6	127	1	PAND_LISIN	Q922a8 listeria in
44	6	4.6	127	1	PAND_LISMO	Q8y603 listeria mo
45	6	4.6	129	1	CU66_HYACE	P45590 hyalophora
46	6	4.6	136	1	RL16_RICPR	Q92c22 rickettsia
47	6	4.6	136	1	RL28_HUMAN	P46779 homo sapien
48	6	4.6	136	1	RL28_MOUSE	P41105 mus musculu
49	6	4.6	136	1	RL28_RAT	P17702 ratu
50	6	4.6	139	1	ADP3_ARATH	Q928K4 arabidopsis
51	6	4.6	139	1	VG39_BPML5	Q05249 mycobacteri
52	6	4.6	145	1	YD54_METUA	Q058749 methanococ
53	6	4.6	148	1	NCCX_ALCXX	Q44582 alcaligenes
54	6	4.6	149	1	VE6_HPV35	P27328 human papil
55	6	4.6	157	1	MOAC_METMA	Q8pxx2 methanosarc
56	6	4.6	163	1	SELT_HUMAN	Q92zj3 homo sapien
57	6	4.6	163	1	URE2_YEREN	P31495 yersinia en
58	6	4.6	166	1	Y483_THERVO	Q97bh2 thermoplas
59	6	4.6	167	1	E1BS_ADE40	P10543 human adeno
60	6	4.6	168	1	PA2F_HUMAN	Q9b2m2 homo sapien
61	6	4.6	170	1	E1BS_ADE41	P10544 human adeno
62	6	4.6	177	1	YDAB_MYCBO	P46830 mycobacteri
63	6	4.6	185	1	EPF_ANAPB	Q44247 anabaena sp
64	6	4.6	186	1	YD66_AQUAE	Q67378 aquifex aeo
65	6	4.6	191	1	YK51_ARCFU	Q28228 archaeoglob
66	6	4.6	195	1	TP15_MOUSE	Q60756 mus musculu
67	6	4.6	196	1	Y01G_HAEIN	P44709 haemophilus
68	6	4.6	197	1	Y01I_METUA	Q60327 methanococ
69	6	4.6	205	1	RUVA_CAUCR	Q9a397 caulobacter
70	6	4.6	208	1	ACPD_STAM	Q932K5 staphylococ
71	6	4.6	208	1	ACPD_STAN	Q99xk1 staphylococ
72	6	4.6	208	1	ACPD_STAM	Q89yk7 staphylococ
73	6	4.6	208	1	GTH1_ARATH	P42760 arabidopsis
74	6	4.6	209	1	GTHB_ARATH	Q928f5 arabidopsis
75	6	4.6	217	1	ENGB_RHIME	Q928f6 rhizobium m
76	6	4.6	227	1	GIDB_MYCPV	Q98x82 mycoplasma
77	6	4.6	227	1	RS3_METKA	Q8x35 methanopyru
78	6	4.6	230	1	TRUB_CYTJO	Q92x36 cytophaga j
79	6	4.6	230	1	MAAI_EMENT	Q43123 emericella
80	6	4.6	232	1	SLT2_CAEBL	Q19892 caenorhabdi
81	6	4.6	235	1	EFPU_BRYPL	P50374 bryopsis pl
82	6	4.6	239	1	DABP_COXBU	P24703 coxiella bu
83	6	4.6	239	1	NEF_HV1SC	Q05857 human immu
84	6	4.6	240	1	Y116_PASMU	Q92cd1 pasteurella
85	6	4.6	243	1	IP1_AGRRH	P14011 agrobacteri
86	6	4.6	244	1	HSBP_HETGL	Q9b19 heterodera
87	6	4.6	248	1	Y4EK_RHISN	P55434 rhizobium s
88	6	4.6	249	1	CXAA_BACTI	Q05069 bacillus th
89	6	4.6	249	1	ECR1_SULSO	Q9uxc4 sulfolobus
90	6	4.6	249	1	VG51_BPT4	P17173 bacterioph
91	6	4.6	249	1	Y0FP_ECOLI	P39298 escherichia
92	6	4.6	252	1	TRP3_COTJA	P06398 coturnix co
93	6	4.6	253	1	TPIS_BACSV	P27876 bacillus su
94	6	4.6	258	1	KLK1_PAPHA	Q28773 papio hamad
95	6	4.6	262	1	KLK1_HUMAN	P12620 gallus gall
96	6	4.6	262	1	TRP3_CHICK	P12620 gallus gall
97	6	4.6	265	1	FABH_ECOLI	P33782 escherichia
98	6	4.6	265	1	RPOD_SULSO	P95889 sulfolobus
99	6	4.6	268	1	YC59_MYCTU	Q11059 mycobacteri
100	6	4.6	274	1	PROC_SOYBN	P17817 glycine max

#### ALIGNMENTS

```

VE6_HPV45
ID VE6_HPV45 STANDARD; PRT; 158 AA.
AC P21735;
DT 01-MAY-1991 (Rel. 18, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE E6 protein.
GN E6.
OS Human papillomavirus type 45.
OC Viruses; dsbDNA viruses, no RNA stage; Papillomaviridae;
OC Papillomavirus.
OX NCBI_TaxID=10593;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94265501; PubMed=8205838;
RA Delius H., Hofmann B.;
RT Curr. Top. Microbiol. Immunol. 186:13-31(1994).
RN [2]
RP SEQUENCE FROM N.A.
RX Kaplan J.B., Burk R.D.;
RL Submitted (Aug-1990) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: THIS PROTEIN HAS TRANSFORMING ACTIVITY IN VITRO.
CC -1- FUNCTION: EXHIBIT A STRONG, BUT NON SPECIFIC AFFINITY FOR DOUBLE
CC STRANDED DNA (IN VITRO).
CC -1- SUBCELLULAR LOCATION: Nuclear matrix-associated.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X74479; CAAS2573.1; -
DR EMBL: M38198; AAA46973.1; -
DR PIR: S36561; S36561.
DR InterPro: IPR001334; E6.
DR Pfam: PF00518; E6; 1.
KM Early protein; DNA-binding; Nuclear protein; zinc-finger;
KW Oncogene.
FT 2N_FING 32 68 POTENTIAL.
FT 2N_FING 105 141
FT 2N_FING 10 10 R -> P (IN REF. 2).
FT 2N_FING 10 10 I -> N (IN REF. 2).
FT 2N_FING 30 30 R -> A (IN REF. 2).
FT 2N_FING 118 118
FT 2N_FING 158 AA; 18897 MW; FICF10DD33AA4C3B CRC64;
SQ SEQUENCE

Query Match 6.9%; Score 9; DB 1; Length 158;
Best Local Similarity 100.0%; Pred. No. 0.096;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 115 GETLEKITT 123
Db 87 GETLEKITT 95

```

```

RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX MEDLINE=96337999; PubMed=8688087;
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
RA Kerlavage A.R., Dougherty B.A., Tomb J.F., Adams M.D., Reich C.I.,
RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhrman J.L., Nguyen D.,
RA Utecherack T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA Corcoran M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA Klenk H.-P., Frazer C.M., Smith H.O., Moese C.R., Venter J.C.;
RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
RT jannaschii."
RL Science 273:1058-1073(1996).
CC -1- SIMILARITY: TO M.JANNASCHII MJ1220 AND MJEC142.
CC -1- SIMILARITY: WITH TYPE I RESTRICTION SYSTEM ADENINE METHYLASES
CC (M SUBUNIT).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U67470; AAB98113.1; -
DR PIR: D64316; D64316.
DR TIGR: MJ0132;
DR InterPro: IPR003356; N6_DNA_Mtase.
DR Pfam: PF02384; N6_Mtase; 1.
KM Hypothetical protein; Complete proteome.
SQ SEQUENCE 220 AA; 25766 MW; 710DDAE4C7A47954 CRC64;

Query Match 6.1%; Score 8; DB 1; Length 220;
Best Local Similarity 100.0%; Pred. No. 1.3;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 102 EXDLIEAI 109
Db 73 EXDLIEAI 80

RESULT 3
DLTC LISIN STANDARD; PRT; 78 AA.
AC 092D49;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DE D-alanine--poly(phosphoribitol)ligase subunit 2 (EC 6.1.1.13) (D-
DE alanyl carrier protein) (DCP).
GN DLTC OR LIN0971.
OS Listeria innocua.
OC Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.
OX NCBI_TaxID=1642;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=CLIP 11262 / Serovar 6a;
RX MEDLINE=21537279; PubMed=11679669;
RA Glaeser P., Frangeul L., Buchrieser C., Ruiniok C., Amend A.,
RA Baquero F., Berche P., Bloecher H., Brandt P., Chakraborty T.,
RA Charbit A., Checouani F., Couve E., de Daruvar A., Dehoux P.,
RA Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,
RA Ertian K.-D., Fiehl H., Garcia-del Portillo F., Garrido P.,
RA Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,
RA Jones L.-M., Kaerst U., Krell T., Kunz M., Kunst F., Kutaypat G.,
RA Madueno E., Maitournam A., Mata Vicente J., Ng E., Nedjati H.,
RA Nardisek G., Novella S., de Pablos B., Perez-Diaz J.-C., Purcell R.,
RA Remmel B., Rose M., Schluter T., Simoes N., Tierrez A.,
RA Vazquez-Boland J.-A., Voss H., Wehland J., Cossart P.;
RL Science 294:849-852(2001).
CC -1- FUNCTION: Involved in the biosynthesis of D-alanyl-lipoteichoic

```

```

CC acid (LTA). Activated D-alanyl-Dcp donates its D-alanyl
CC substituent to membrane-associated LTA (By similarity).
CC -1- CATALYTIC ACTIVITY: ATP + D-alanine + poly(ribitol phosphate) =
CC AMP + diphosphate + O-D-alanyl-poly(ribitol phosphate).
CC -1- PATHWAY: D-alanyl-lipoteichoic acid biosynthesis.
CC -1- PPM: 4'-phosphopantetheine is transferred from CoA to a specific
CC serine of apo-Dcp (By similarity).
CC -1- SIMILARITY: Contains 1 acyl carrier domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sb-sib.ch).
CC -----
CC EMBL: A1596167; CAC96202.1; -.
CC PIR: A15154; A15154.
CC L181186; L180971; -.
CC HAMAP: MF_00565; -.
CC InterPro: IPR003230; D-ala carrier.
CC InterPro: IPR006163; Pp_bind.
CC PROSITE: PD015103; D-ala carrier; 1.
CC PROSITE: PS50075; ACP DOMAIN; FALSE NEG.
CC Ligase; Cell wall; Phosphopantetheine; Complete proteome.
CC BINDING 36 36 PHOSPHOPANTHETHEINE (PROBABLE).
CC SEQUENCE 78 AA; 8996 MW; EBE69446CCB96DBD CRC64;

Query Match 5.3%; Score 7; DB 1; Length 78;
Best Local Similarity 100.0%; Pred. No. 5.5;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 69 IIEISRL 75
DB 44 IIEISRL 50

```

```

CC -1- SUBUNIT: F-TYPE ATPASES HAVE 2 COMPONENTS, CF(1) - THE CATALYTIC
CC CORE - AND CF(0) - THE MEMBRANE PROTON CHANNEL. CF(1) HAS FIVE
CC SUBUNITS: ALPHA(3), BETA(3), GAMMA(1), DELTA(1), EPSILON(1). CF(0)
CC HAS THREE MAIN SUBUNITS: A, B AND C.
CC -1- SIMILARITY: BELONGS TO THE ATPASE C CHAIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sb-sib.ch).
CC -----
CC EMBL: D16466; BAA03937.1; -.
CC EMBL: AB000109; BAA78065.1; -.
CC PIR: S68159; S68159.
CC DictyDb: DD07777; atp9.
CC InterPro: IPR002379; ATPase_Csub.
CC InterPro: IPR000454; Sub ATPase_Csub.
CC Pfam: PF00137; ATP-synt C; 1.
CC PRINTS: PR00124; ATPaseC.
CC PROSITE: PS00605; ATPase_C; FALSE NEG.
CC Hydrogen ion transport; Lipid-binding; Mitochondrion; CF(0);
CC Transmembrane.
CC SEQUENCE 88 AA; 8995 MW; F52709F8FB05B53A CRC64;

Query Match 5.3%; Score 7; DB 1; Length 88;
Best Local Similarity 100.0%; Pred. No. 6.1;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 56 LELASAV 62
DB 10 LELASAV 16

```

```

RESULT 4
ATP9_DICDI STANDARD; PRT; 88 AA.
ID ATP9_DICDI STANDARD; PRT; 88 AA.
AC Q37315;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE ATP synthase protein 9, mitochondrial (Ec 3.6.3.14) (Lipid-binding,
DE protein).
OS Dictyostelium discoideum (Slime mold).
OC Eukaryota; Eukaryota; Dictyostelida; Dictyostelium.
OC NCBI_TaxID=44689;
OX NCBI_TaxID=44689;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AX3;
RX MEDLINE=95254668; PubMed=7736610;
RA Angata K., Kuroe K., Yanagisawa K., Tanaka Y.;
RT "Codon usage, genetic code and phylogeny of Dictyostelium discoideum
RT mitochondrial DNA as deduced from a 7.3-Kb region.";
RL Curr. Genet. 27:249-256(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=AX3;
RX MEDLINE=20279206; PubMed=10821186;
RA Ogawa S., Yoshino R., Angata K., Iwamoto M., Pi M., Kuroe K.,
RA Matsuo K., Morio T., Urushihara H., Yanagisawa K., Tanaka Y.;
RT "The mitochondrial DNA of Dictyostelium discoideum: complete sequence,
RT gene content and genome organization.";
RL Mol. Gen. Genet. 263:514-519(2000).
CC -1- FUNCTION: THIS PROTEIN IS ONE OF THE CHAINS OF THE NONENZYMATIC
CC MEMBRANE COMPONENT (F0) OF MITOCHONDRIAL ATPase.
CC -1- CATALYTIC ACTIVITY: ATP + H(2)O + H(+) (In) = ADP + phosphate +
CC H(+) (Out).

```

```

RESULT 5
R18E_PYPAA STANDARD; PRT; 122 AA.
ID R18E_PYPAA STANDARD; PRT; 122 AA.
AC O82Y02;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE 50S ribosomal protein L18e.
GN RPL18E OR PAB0672.
OS Pyrobaculum aerophilum.
OC Archaea; Crenarchaeota; Thermoprotei; Thermoproteales;
OC Thermoproteaceae; Pyrobaculum.
OX NCBI_TaxID=13773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=IM2 / ATCC 51768 / DSM 7523;
RX MEDLINE=21664397; PubMed=11792869;
RA Fitz-Gibbon S.T., Ladner H., Kim U.-J., Stetter K.O., Simon M.I.,
RA Miller J.H.;
RT "Genome sequence of the hyperthermophilic crenarchaeon Pyrobaculum
RT aerophilum";
RL Proc. Natl. Acad. Sci. U.S.A. 99:984-989(2002).
CC -1- SIMILARITY: BELONGS TO THE L18E FAMILY OF RIBOSOMAL PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sb-sib.ch).
CC -----
CC EMBL: AE009779; AAL62941.1; -.
CC HAMAP: MF_00329; -.
CC InterPro: IPR000039; Ribosomal_L18e.
CC Pfam: PF00256; L15; 1.
CC PROSITE: PS01106; RIBOSOMAL_L18E; FALSE NEG.

```

KW Ribosomal protein; Complete proteome.  
 SQ SEQUENCE 122 AA; 13252 MW; 88F3DB732CAE0394 CRC64;  
 Query Match 5.3%; Score 7; DB 1; Length 122;  
 Best Local Similarity 100.0%; Pred. No. 8.1;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 36 GSGGRIV 42  
 DB 115 GSGGRIV 121

RESULT 6  
 ID Y964 MYCTU STANDARD; PRT; 160 AA.  
 AC P715T6;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Hypothetical protein RV0964C.  
 GN RV0964C OR MT0992.1 OR MTCY10D7.10.  
 OS Mycobacterium tuberculosis.  
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
 OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
 OX NCBI\_TaxID=1773;  
 [1]  
 SEQUENCE FROM N.A.  
 RP SRRAIN=H37Rv;  
 RC MEDLINE=98295987; PubMed=9634230;  
 RX Cole S.T., Broesch R., Parkhill J., Garnier T., Churcher C., Harris D.,  
 RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,  
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,  
 RA Davies R., Devlin K., Felwell T., Gentles S., Hamlin N., Holroyd S.,  
 RA Hornaby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,  
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers R.,  
 RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,  
 RA Sultson J.E., Taylor K., Whitehead S., Barrell B.G.;  
 RA "Deciphering the biology of Mycobacterium tuberculosis from the  
 RT complete genome sequence.";  
 RL Nature 393:537-544(1998).  
 [2]  
 SEQUENCE FROM N.A.  
 RP SRRAIN=CDC 1551 / Oshkosh;  
 RC Fleishmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,  
 RA Peterson J., Deboy R., Dodson R., Gwinn M.L., Hatt D., Hickey E.,  
 RA Kolonay J.F., Nelson W.C., Umayan L.A., Ermolova M.D., Salzberg S.L.,  
 RA Delcher A., Ustebek T., Weidman J., Khouri H., Gill J., Mikula A.,  
 RA Bishai W.;  
 RT "Whole genome comparison of Mycobacterium tuberculosis clinical and  
 RT laboratory strains.";  
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.

CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----  
 CC EMBL; Z79700; CAB01987.1; -  
 DR EMBL; AE006984; AAK45241.1; ALT\_INIT.  
 DR PIR; B70718; B70718.  
 DR TTGR; MT0992.1; -  
 DR Tubercular; RV0964C; -  
 KM Hypothetical protein; Complete proteome.  
 FT CONFLICT 124 P -> T (IN REF. 2).  
 FT NP\_BIND 124 P -> T (IN REF. 2).  
 SQ SEQUENCE 160 AA; 17317 MW; 2B6EC9285852211F CRC64;

Query Match 5.3%; Score 7; DB 1; Length 160;  
 Best Local Similarity 100.0%; Pred. No. 10;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 108 AIRRASN 114  
 DB 95 AIRRASN 101

RESULT 7  
 ID ENGB AGRT5 STANDARD; PRT; 217 AA.  
 AC CEUB4;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Probable gtp-binding protein engb.  
 GN ENGB OR ATU0383 OR AGS C.672.  
 OS Agrobacterium tumefaciens (strain C58 / ATCC 33970).  
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;  
 OC Rhizobiaceae; Rhizobium/Agrobacterium group; Agrobacterium.  
 OX NCBI\_TaxID=176299;  
 [1]  
 SEQUENCE FROM N.A.  
 RP MEDLINE=21608550; PubMed=11743193;  
 RX Wood D.W., Serubal J.C., Kaul R., Monks D.E., Kitaajima J.P., Woo L.,  
 RA Okura V.K., Zhou Y., Chen L., Wood G.E., Almeida N.F. Jr., Woo L.,  
 RA Chen Y., Paulsen I.T., Eisen J.A., Karp P.D., Bovee D. Sr., C.,  
 RA Chapman P., Clendenning J., Deatherage G., Gillet W., Grant C.,  
 RA Kutayavin T., Levy R., Li M.-J., McClelland E., Palmeri A., Gordon D.,  
 RA Raymond C., Rouse G., Saenphimachak C., Wu Z., Romero P., Perry M.,  
 RA Zhang S., Yoo H., Tao Y., Biddle P., Jung M., Krespan W., Dolan M.,  
 RA Gordon-Kamm B., Liao L., Kim S., Hendrick C., Zhao Z.-Y., Dolan M.,  
 RA Chumley F., Tingey S.V., Tomb J.-F., Gordon M.P., Olson M.V.,  
 RA Nester E.W.;  
 RT "The genome of the natural genetic engineer Agrobacterium tumefaciens  
 RT C58.";  
 RL Science 294:2317-2323(2001).  
 [2]  
 SEQUENCE FROM N.A.  
 RP MEDLINE=21608551; PubMed=11743194;  
 RX Goodner B., Hinkle G., Galtung S., Miller N., Blanchard M.,  
 RA Guorillo B., Goldman B.S., Cao Y., Askenazi M., Halling C., Mullin L.,  
 RA Hummel K., Gordon J., Vaudin M., Iarchouk O., Epp A., Liu F.,  
 RA Woliam C., Allinger M., Doughty D., Scott C., Leppa C., Markelz B.,  
 RA Flanagan C., Crowell C., Gurson J., Lomo C., Sear C., Strub G.,  
 RA Cielo C., Slater S.;  
 RT "Genome sequence of the plant pathogen and biotechnology agent  
 RT Agrobacterium tumefaciens C58.";  
 RL Science 294:2323-2328(2001).

CC -----  
 CC -FUNCTION: Necessary for normal cell division and for the  
 CC maintenance of normal septation (by similarity).  
 CC -SIMILARITY: BELONGS TO THE ENGB FAMILY.

CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----  
 CC EMBL; AE009009; AAL41405.1; -  
 DR EMBL; AE007975; AAK86199.1; -  
 DR PIR; AG2623; AG2623.  
 DR PIR; P97405; P97405.  
 DR HAMAP; MF 00321; -1.  
 DR InterPro; IPR005289; GTP-binding\_dom.  
 DR TIGRFAMs; TIGR00650; MG442; 2.  
 KM Cell division; Septation; GTP-binding; Complete proteome.  
 FT NP\_BIND 41 48 GTP (POTENTIAL).  
 FT NP\_BIND 95 99 GTP (POTENTIAL).  
 FT NP\_BIND 162 165 GTP (POTENTIAL).  
 SQ SEQUENCE 217 AA; 23864 MW; B4E7B9605ABE192 CRC64;

Query Match 5.3%; Score 7; DB 1; Length 217;  
 Best Local Similarity 100.0%; Pred. No. 13;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 115 GETLEKI 121  
 |||||  
 Db 177 GETLEKI 183

RESULT 8  
 TEVI\_BPT4 STANDARD; PRT; 245 AA.  
 ID TEVI\_BPT4  
 AC P13299; O97073;  
 DT 01-JAN-1990 (Rel. 13, Created)  
 DT 15-JUL-1999 (Rel. 38, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE Intron-associated endonuclease 1 (EC 3.1.-.-) (I-TEVI) (IRF protein).  
 GN ITEVIR.  
 OS Bacteriophage T4.  
 OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Myoviridae;  
 OC T4-like viruses.  
 OX NCBI\_TaxID=10665;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=86189927; PubMed=3698096;  
 RA Chu F.K., Maley G.F., West D.K., Belfort M., Maley F.;  
 RT "Characterization of the intron in the phage T4 thymidylate synthase  
 gene and evidence for its self-excision from the primary  
 transcript.";  
 RT Cell 45:157-166 (1986).  
 RL (2)  
 RN REVISION TO 67.  
 RC STRAIN=ALC4;  
 RA Chu F.;  
 RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.  
 RN (3)  
 RP SEQUENCE FROM N.A.  
 RA Kutter E., Arisaka F., Kunisawa T., Tsugita A., Mosig G.,  
 Wesenitzlinov V., Ruger W., Stidham T., Thomas E.;  
 RT "Bacteriophage T4 genome analysis.";  
 RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.  
 RN (4)  
 RP IDENTIFICATION OF PROTEIN  
 RX MEDLINE=89278087; PubMed=2543665;  
 RA West D.K., Changchien L.-M., Maley G.F., Maley F.;  
 RT "Evidence that the intron open reading frame of the phage T4 td gene  
 encodes a specific endonuclease.";  
 RL J. Biol. Chem. 264:10343-10346 (1989).  
 CC -1- FUNCTION: THIS ENDONUCLEASE IS SPECIFIC TO THE THYMIDYLATE  
 SYNTHASE (TD) GENE SPLICING AND IS INVOLVED IN INTRON  
 REMOVAL.  
 CC -1- COFACTOR: MAGNESIUM.  
 CC -1- SIMILARITY: TO ENDONUCLEASES OF GROUP I INTRONS OF FUNGI AND  
 PHAGE.

-----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

-----  
 CC EMBL; M12742; AAC12817.1; -  
 CC EMBL; AF158101; AAD42521.2; ALT\_INIT.  
 CC PIR; T10130; T10130.  
 CC PDB; 1T3J; 13-JUL-01.  
 CC PDB; 1LNO; 30-OCT-02.  
 CC PDB; 1MK0; 30-OCT-02.  
 CC REBASE; 2625; 1-TEVI.  
 CC InterPro; IPR006350; Intron\_endog1.  
 CC InterPro; IPR003611; Intron\_nuc\_2.  
 CC InterPro; IPR000305; UYRC\_N.  
 CC Pfam; PF01541; Excl\_endo\_N; 1.  
 CC SMART; SM00465; G1YC; 1.

DR SMART; SM00496; IENR2; 2.  
 DR TIGRFA; TIGR01453; gryp1ntron\_endo; 1.  
 DR PROSITE; PS50164; UYRC\_1; 1.  
 KW Hydrolase; Nuclease; Endonuclease; Intron homing; Magnesium;  
 KW 3d-structure.  
 SQ SEQUENCE 245 AA; 28232 MW; 29DF2ACF08EC64C CRC64;

Query Match 5.3%; Score 7; DB 1; Length 245;  
 Best Local Similarity 100.0%; Pred. No. 15;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 100 PYEKDII 106  
 |||||  
 Db 65 PYEKDII 71

RESULT 9  
 PMT2\_TOBAC STANDARD; PRT; 353 AA.  
 ID PMT2\_TOBAC  
 AC O9SEH7;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Putrescine N-methyltransferase 2 (EC 2.1.1.53) (PMT 2).  
 GN PMT2.  
 OS Nicotiana tabacum (Common tobacco).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 OC Asteridae; Lamiales; Solanales; Solanaceae; Nicotiana.  
 OX NCBI\_TaxID=4097;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. Xanthi;  
 RX MEDLINE=20064975; PubMed=10598105;  
 RA Riechers D.B., Timko W.P.;  
 RT "Structure and expression of the gene family encoding putrescine N-  
 methyltransferase in Nicotiana tabacum: new clues to the evolutionary  
 origin of cultivated tobacco.";  
 RL Plant Mol. Biol. 41:387-401 (1999).  
 CC -1- CATALYTIC ACTIVITY: S-adenosyl-L-methionine + putrescine = S-  
 CC adenosyl-L-homocysteine + N-methylputrescine.  
 CC -1- PATHWAY: Nicotine biosynthesis.  
 CC -1- SIMILARITY: BELONGS TO THE SPERMIDINE/SPERMINE SYNTHASES FAMILY.

-----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

-----  
 CC EMBL; AF126809; AAF14878.1; -  
 CC InterPro; IPR000051; SAM bind.  
 CC InterPro; IPR001045; Spermine synthase.  
 CC Pfam; PF01564; Spermine synth; 1.  
 CC TIGRFAMs; TIGR00417; spse; 1.  
 DR PROSITE; PS01330; SPERMIDINE\_SYNTHASE; 1.  
 DR TRANSFERASE; Methyltransferase; Multigene family.  
 KW SEQUENCE 353 AA; 38746 MW; DD2E024121D511D CRC64;

Query Match 5.3%; Score 7; DB 1; Length 353;  
 Best Local Similarity 100.0%; Pred. No. 20;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 95 ENGGEPPY 101  
 |||||  
 Db 119 ENGGEPPY 125

RESULT 10  
 PMT1\_TOBAC STANDARD; PRT; 375 AA.  
 ID PMT1\_TOBAC

```

AC Q42963; Q9SEH6;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Putrescine N-methyltransferase 1 (EC 2.1.1.53) (PMT 1) (A411).
GN PMT.
OS Nicotiana tabacum (Common tobacco).
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; Lamiales; Solanales; Solanaceae; Nicotiana.
OC NCBI_TaxID=4097;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Burley 21; TISSUE=Root;
RX MEDLINE=94312878; PubMed=8038607;
RA Hibl N., Higashiguchi S., Hashimoto T., Yamada Y.;
RT "Gene expression in tobacco low-nicotine mutants.";
RL Plant Cell 6:723-735(1994).
(2)
RN SEQUENCE FROM N.A.
RC STRAIN=cv. Xanthi;
RX MEDLINE=20064975; PubMed=10598105;
RA Riechers D.E., Timko M.P.;
RT "Structure and expression of the gene family encoding putrescine N-
RT methyltransferase in Nicotiana tabacum: new clues to the evolutionary
RT origin of cultivated tobacco.";
RL Plant Mol. Biol. 41:387-401(1999).
CC -1- CATALYTIC ACTIVITY: S-adenosyl-L-methionine + putrescine = S-
CC adenosyl-L-homocysteine + N-methylputrescine.
CC -1- PATHWAY: Nicotinic biosynthesis.
CC -1- TISSUE SPECIFICITY: PREDOMINANTLY EXPRESSED IN ROOTS.
CC -1- SIMILARITY: BELONGS TO THE SPERMIDINE/SPERMINE SYNTHASE FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL: D28506; BAA05867.1; -
CC EMBL: AF126810; AAF14879.1; -
DR PTR: T03681; T03681.
DR InterPro: IPR000051; SAM_bind.
DR InterPro: IPR001045; Spermine synthase.
DR Pfam: PF01564; Spermine synth. 1.
DR TIGRPFAMs: TIGR00417; SPEE. 1.
DR PROSITE: PS01330; SPERMIDINE SYNTHASE; 1.
KW Transferase; Methyltransferase; Multigene family.
FT CONFLICT 308 308 A -> V (IN REF. 2).
FT SEQUENCE 375 AA; 4113 MW; 9E6483CDD4371287 CRC64;
SQ
Query Match 5.3%; Score 7; DB 1; Length 375;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 95 ENGPFY 101
Db 141 ENGPFY 147
RESULT 11
PRLP RAT STANDARD; PRT; 377 AA.
AC Q9ROPS;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Prolagin precursor (Proline-arginine-rich end leucine-rich repeat
DE protein).
GN PRELP.
OS Rattus norvegicus (Rat).

```

```

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OC NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Connective tissue;
RX MEDLINE=20576219; PubMed=11007795;
RA Bengtsson E., Aspegren A., Heinegaard D., Sommer Y., Spillmann D.;
RT "The amino-terminal part of PRELP binds to heparin and heparan
RT sulfate.";
RL J. Biol. Chem. 275:40695-40702(2000).
CC -1- FUNCTION: May anchor basement membranes to the underlying
CC connective tissue (By similarity).
CC -1- SUBUNIT: Binds the basement membrane heparan sulfate proteoglycan
CC perlecan and triple helical collagen type I and type II (By
CC similarity).
CC -1- SUBCELLULAR LOCATION: Secreted; extracellular matrix.
CC -1- DOMAIN: The basic amino-terminal Arg/Pro-rich binds heparin and
CC heparan sulfate. Binds collagens type I and type II through its
CC leucine-rich repeat domain (By similarity).
CC -1- SIMILARITY: BELONGS TO THE SMALL LEUCINE-RICH PROTEOGLYCAN
CC (SLRP) FAMILY CLASS II SUBFAMILY.
CC -1- SIMILARITY: Contains 12 leucine-rich (LRR) repeats.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL: AF163569; AAG23724.1; -
DR InterPro: IPR001611; LRR.
DR InterPro: IPR000372; LRR_Nterm.
DR InterPro: IPR003591; LRR_LYP.
DR Pfam: PF00560; LRR; 7.
DR Pfam: PF01462; LRRNT; 1.
DR SMART: SM00013; LRRNT; 1.
KW Glycoprotein; Extracellular matrix; Repeat; Leucine-rich repeat;
KW Signal.
FT SIGNAL 1 21
FT CHAIN 22 377
FT DOMAIN 68 84
FT REPEAT 90 109
FT REPEAT 110 133
FT REPEAT 134 157
FT REPEAT 158 178
FT REPEAT 179 202
FT REPEAT 203 228
FT REPEAT 229 249
FT REPEAT 250 273
FT REPEAT 274 298
FT REPEAT 299 318
FT REPEAT 319 357
FT REPEAT 358 377
FT DOMAIN 192 197
FT DISULFID 327 368
FT CARBOHYD 119 119
FT CARBOHYD 284 284
FT CARBOHYD 315 315
FT CARBOHYD 322 322
SQ SEQUENCE 377 AA; 43179 MW; 79CBB62534753C46 CRC64;
Query Match 5.3%; Score 7; DB 1; Length 377;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 90 VFSKLEN 96
Db 185 VFSKLEN 191

```

RESULT 12  
 PRLP\_MOUSE STANDARD; PRT; 378 AA.  
 AC Q9JK53;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE Prolargin precursor (Proline-arginine-rich and leucine-rich repeat protein).  
 DE PRLP.  
 GN Mus musculus (Mouse).  
 OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 NCBI\_TaxId=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=21588626; PubMed=11731272;  
 RA Grover J, Roughley P.J.  
 RL "Characterization and expression of murine PRLP.",  
 Matrix Biol. 20:555-564(2001).  
 [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Salivary gland;  
 RX MEDLINE=22388257; PubMed=12477932;  
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
 Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
 Diatchenko L., Matsushita K., Farmer A.A., Rubin G.M., Hong L.,  
 Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
 Brownstein M.J., Ushed T.B., Tomihata S., Carninci P., Prange C.,  
 Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mallaby S.J.,  
 Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hultky S.W.,  
 Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 Fahey J., Helton E., Kettner M., Madan A., Rodriguez S., Sanchez A.,  
 Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
 Butcherfield V.S.N., Krzywinski M.I., Skalska U., Smalls D.E.,  
 Scherch A., Schein J.E., Jones S.J.M., Marra M.A.,  
 "Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences.",  
 Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 RL  
 CC -1- FUNCTION: May anchor basement membranes to the underlying connective tissue (By similarity).  
 CC -1- SUBUNIT: Binds the basement membrane heparan sulfate proteoglycan perlecan and triple helical collagens type I and type II (By similarity).  
 CC -1- SUBCELLULAR LOCATION: Secreted; extracellular matrix.  
 CC -1- TISSUE SPECIFICITY: Expressed in cartilage throughout both fetal development and post-natal life. It is also expressed in the developing embryo prior to skeletogenesis. In adult, highest expression in lung, lower levels in cardiac and skeletal muscle.  
 CC -1- DOMAIN: The basic amino-terminal Arg/Pro-rich binds heparin and heparan sulfate. Binds collagens type I and type II through its leucine-rich repeat domain (By similarity).  
 CC -1- SIMILARITY: BELONGS TO THE SMALL LEUCINE-RICH PROTEOGLYCAN (SLRP) FAMILY. CLASS II SUBFAMILY.  
 CC -1- SIMILARITY: Contains 12 leucine-rich (LRR) repeats.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; AF261888; AAF72994.2; -  
 CC EMBL; AF261887; AAF72994.2; JOINED.  
 CC EMBL; BC019775; AAH19775.1; -  
 CC MGD; MGI:2151110; Prlp.

DR InterPro; IPR001611; LRR.  
 DR InterPro; IPR000372; LRR\_N-term.  
 DR InterPro; IPR003591; LRR\_Typ.  
 DR Pfam; PF00560; LRR; 7.  
 DR Pfam; PF01462; LRRNT; 1.  
 DR SMART; SM00013; LRRNT; 1.  
 KW Glycoprotein; Extracellular matrix; Repeat; Leucine-rich repeat; Signal.  
 FT SIGNAL. 1 21 POTENTIAL.  
 FT CHAIN 22 378 PROLARGIN.  
 FT DOMAIN 69 85 CIS-RICH.  
 FT REPEAT 91 110 LRR-S 1.  
 FT REPEAT 111 134 LRR-T 1.  
 FT REPEAT 135 158 LRR-T 2.  
 FT REPEAT 159 179 LRR-S 2.  
 FT REPEAT 180 203 LRR-T 3.  
 FT REPEAT 204 229 LRR-T 4.  
 FT REPEAT 230 250 LRR-S 3.  
 FT REPEAT 251 274 LRR-T 5.  
 FT REPEAT 275 299 LRR-T 6.  
 FT REPEAT 300 319 LRR-S 4.  
 FT REPEAT 320 358 LRR-T 7.  
 FT REPEAT 359 378 LRR-T 8.  
 FT DOMAIN 133 198 POLY-LEU.  
 FT DISULFID 328 369 BY SIMILARITY.  
 FT CARBOHYD 120 120 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 285 285 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 316 316 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 323 323 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 378 AA; 43292 MW; 81654FBD5F55186 CRC64;

Query Match 5.3%; Score 7; DB 1; Length 378;  
 Best Local Similarity 100.0%; Pred. No. 22;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 90 VFESKLEN 96  
 Db 186 VFESKLEN 192

RESULT 13  
 PWT3\_TOBAC STANDARD; PRT; 381 AA.  
 AC Q9SEH5;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Putrescine N-methyltransferase 3 (EC 2.1.1.53) (PWT 3).  
 GN PWT3.  
 OS Nicotiana tabacum (Common tobacco).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 OC Asterales; Lamiales; Solanales; Solanaceae; Nicotiana.  
 NCBI\_TaxId=4097;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. Xanthi;  
 RX MEDLINE=20064975; PubMed=10598105;  
 RA Riechers D.E., Timko M.P.,  
 "Structure and expression of the gene family encoding putrescine N-methyltransferase in Nicotiana tabacum: new clues to the evolutionary origin of cultivated tobacco.",  
 Plant Mol. Biol. 41:387-401(1999).  
 RL  
 CC -1- CATALYTIC ACTIVITY: S-adenosyl-L-methionine + putrescine = S-adenosyl-L-homocysteine + N-methylputrescine.  
 CC -1- PATHWAY: Nicotine biosynthesis.  
 CC -1- SIMILARITY: BELONGS TO THE SPERMIDINE/SPERMINE SYNTHASIS FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial



entitities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; AF126811; AAF14880.1; -

CC InterPro; IPR000051; SAM bind.

DR InterPro; IPR001045; Spermine synthase.

DR Pfam; PF01564; Spermine\_synth; 1.

DR TIGRfam; TIGR00417; sps; 1.

DR PROSITE; PS01330; SPERMIDINE\_SYNTHASE; 1.

DR Transferase; Methyltransferase; Multigene family.

SO SEQUENCE 381 AA; 41795 MW; A19DE2B83CE4D530 CRC64;

Query Match 5.3%; Score 7; DB 1; Length 381;

Best Local Similarity 100.0%; Pred. No. 22;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 95 ENGPFY 101

Db 147 ENGPFY 153

RESULT 14

PRLP\_BOVIN STANDARD; PRT; 381 AA.

AC Q9GK8; 28-FEB-2003 (Rel. 41, Created)

DT 28-FEB-2003 (Rel. 41, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Prolargin precursor (Proline-arginine-rich end leucine-rich repeat protein).

GN PRELP.

OS Bos taurus (Bovine).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovinae; Bovinae; Bos.

OX NCBI\_TaxID=9913;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Articular cartilage;

RX MEDLINE=20576219; PubMed=11007795;

RA Bengtsson E., Aspberg A., Heinegaard D., Sommarin Y., Spillmann D.;

RT "The amino-terminal part of PRELP binds to heparin and heparan sulfate."

RT J. Biol. Chem. 275:40695-40702(2000).

RL [2]

RN FUNCTION.

RX MEDLINE=1964083; PubMed=11847210;

RA Bengtsson E., Moergelin M., Sasaki T., Timpl R., Heinegaard D., Aspberg A.;

RT "The leucine-rich repeat protein PRELP binds perlecan and collagens and may function as a basement membrane anchor."

RT J. Biol. Chem. 277:15061-15068(2002).

CC -1- FUNCTION: May anchor basement membranes to the underlying connective tissue.

CC -1- SUBUNIT: Binds the basement membrane heparan sulfate proteoglycan perlecan and triple helical collagens type I and type II.

CC -1- SUBCELLULAR LOCATION: Secreted; extracellular matrix.

CC -1- DOMAIN: The basic amino-terminal Arg/Pro-rich binds heparan and heparan sulfate. Binds collagens type I and type II through its leucine-rich repeat domain.

CC -1- SIMILARITY: BELONGS TO THE SMALL LEUCINE-RICH PROTEOGLYCAN (SLRP) FAMILY. CLASS II SUBFAMILY.

CC -1- SIMILARITY: Contains 12 leucine-rich (LRR) repeats.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; AF163568; AAG3723.1; -

DR InterPro; IPR001611; LRR.

DR InterPro; IPR000372; LRR\_Nterm.

DR InterPro; IPR003591; LRR\_Cyp.

DR Pfam; PF00560; LRR; 9.

DR Pfam; PF01462; LRRNT; 1.

DR PRINTS; PR00019; LEURICHRPT.

DR SMART; SM00013; LRRNT; 1.

KW Glycoprotein; Extracellular matrix; Repeat; Leucine-rich repeat; Signal.

FT SIGNAL 1 21

FT CHAIN 22 381

FT DOMAIN 72 88

FT REPEAT 94 113

FT REPEAT 114 137

FT REPEAT 138 161

FT REPEAT 162 182

FT REPEAT 183 206

FT REPEAT 207 232

FT REPEAT 233 253

FT REPEAT 254 277

FT REPEAT 278 302

FT REPEAT 303 322

FT REPEAT 323 361

FT REPEAT 362 381

FT DOMAIN 196 201

FT DISULFID 331 372

FT CARBOHYD 123 123

FT CARBOHYD 288 288

FT CARBOHYD 319 319

FT CARBOHYD 326 326

SO SEQUENCE 381 AA; 43682 MW; 23DA99C01BB772A0 CRC64;

Query Match 5.3%; Score 7; DB 1; Length 381;

Best Local Similarity 100.0%; Pred. No. 22;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 90 VFSEKLEN 96

Db 189 VFSEKLEN 195

RESULT 15

PRLP\_HUMAN STANDARD; PRT; 382 AA.

AC P51888;

DT 01-OCT-1996 (Rel. 34, Created)

DT 01-OCT-1996 (Rel. 34, Last sequence update)

DT 15-SEP-2003 (Rel. 42, Last annotation update)

DE Prolargin precursor (Proline-arginine-rich end leucine-rich repeat protein).

GN PRELP.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

OX NCBI\_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A. AND PARTIAL SEQUENCE.

RX MEDLINE=96029653; PubMed=7592739;

RA Bengtsson E., Neame P.J., Heinegaard D., Sommarin Y.;

RT "The primary structure of a basic leucine-rich repeat protein, PRELP, found in connective tissues."

RT J. Biol. Chem. 270:25639-25644(1995).

RL [2]

RN SEQUENCE FROM N.A.

RX MEDLINE=9127584; PubMed=8954791;

RA Grover J., Chen X.-N., Korenberg J.R., Recklies A.D., Roughley P.J.;

RT "The gene organization, chromosome location, and expression of a 55-kDa matrix protein (PRELP) of human articular cartilage."

RL Genomics 38:109-117(1996).

RN [3]

RP SEQUENCE FROM N.A.

RC TISSUE=Pancreas; and Spleen;

RX MEDLINE=22388257; PubMed=12477932;



RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shennan C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heish F.,  
 RA Datchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Udén T.B., Toshiyuki S., Carninci P., Prange C.,  
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mulhally S.J.,  
 RA Bosak S.A., Mcwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 RA Vallian D.K., Munz D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahy J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalins D.E.,  
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.,  
 RT "Generation and initial analysis of more than 15,000 full-length  
 RT human and mouse cDNA sequences.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 CC -1- FUNCTION: May anchor basement membranes to the underlying  
 CC connective tissue (By similarity).  
 CC -1- SUBUNIT: Binds the basement membrane heparan sulfate proteoglycan  
 CC perlecan and triple helical collagens type I and type II (By  
 CC similarity).  
 CC -1- SUBCELLULAR LOCATION: Secreted; extracellular matrix.  
 CC -1- TISSUE SPECIFICITY: Connective tissue.  
 CC -1- DOMAIN: The basic amino-terminal Arg/Pro-rich binds heparin and  
 CC heparan sulfate. Binds collagens type I and type II through its  
 CC leucine-rich repeat domain (By similarity).  
 CC -1- SIMILARITY: BELONGS TO THE SMALL LEUCINE-RICH PROTEOGLYCAN  
 CC (SLRP) FAMILY. CLASS II SUBFAMILY.  
 CC -1- SIMILARITY: Contains 12 leucine-rich (LRR) repeats.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL, U29089; AAC0230.1; -;  
 DR EMBL, U41344; AAC18782.1; -;  
 DR EMBL, U41343; AAC18782.1; JOINED.  
 DR EMBL, BC032498; AAH32498.1; -;  
 DR PIR, I39068; I39068.  
 DR Genew; HGNC:9357; PRLP.  
 DR MIM; 601914; -;  
 DR GO; GO:0005578; C:extracellular matrix; TMS.  
 DR GO; GO:0005201; P:extracellular matrix structural constituent; TMS.  
 DR GO; GO:0001501; P:skeletal development; TMS.  
 DR InterPro; IPR001611; LRR.  
 DR InterPro; IPR000372; LRR\_Nterm.  
 DR Pfam; PF00560; LRR; 8.  
 DR Pfam; PF01462; LRRNT; 1.  
 DR SMART; SM00013; LRRNT; 1.  
 DR KW Glycoprotein; Extracellular matrix; Repeat; Leucine-rich repeat;  
 KM Signal; Polymorphism.  
 FT SIGNAL 1 20  
 FT CHAIN 1 382  
 FT DOMAIN 73 89  
 FT REPEAT 95 114 LRR-S 1.  
 FT REPEAT 115 138 LRR-T 1.  
 FT REPEAT 139 162 LRR-T 2.  
 FT REPEAT 163 183 LRR-S 2.  
 FT REPEAT 184 207 LRR-T 3.  
 FT REPEAT 208 233 LRR-T 4.  
 FT REPEAT 234 254 LRR-S 3.  
 FT REPEAT 255 278 LRR-T 5.  
 FT REPEAT 279 303 LRR-T 6.  
 FT REPEAT 304 323 LRR-S 4.  
 FT REPEAT 324 362 LRR-T 7.

FT REPEAT 363 382 LRR-T 8.  
 FT DOMAIN 197 202 POLY-LEU.  
 FT DISULFID 332 373 BY SIMILARITY.  
 FT CARBOHYD 124 124 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT CARBOHYD 289 289 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT CARBOHYD 320 320 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT CARBOHYD 327 327 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT VARIANT 348 348 N -> H (IN dbSNP:9439).  
 FT SEQUENCE 382 AA; 43809 MW; AICAE16B7515695 CRC64;  
 SQ  
 Query Match 5.3%; Score 7; DB 1; Length 382;  
 Best Local Similarity 100.0%; Pred. No. 22;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 90 VFVKLEN 96  
 DB 190 VFVKLEN 196  
 RESULT 16  
 ID PMT4\_T0BAC STANDARD; PRT; 419 AA.  
 AC Q9SER4;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Putrescine N-methyltransferase 4 (EC 2.1.1.53) (PMT 4).  
 GN PMT4.  
 OS Nicotiana tabacum (Common tobacco).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 OC Asteridae; lamiales; Solanales; Solanaceae; Nicotiana.  
 OX NCBI\_TaxId=4097;  
 RN [1] TaxId=4097;  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. Xanthi.  
 RX MEDLINE=20064975; PubMed=10598105;  
 RA Riechers D.E., Timko M.P.;  
 RT "Structure and expression of the gene family encoding putrescine N-  
 RT methyltransferase in Nicotiana tabacum: new clues to the evolutionary  
 RT origin of cultivated tobacco.";  
 RL Plant Mol. Biol. 41:387-401(1999).  
 CC -1- CATALYTIC ACTIVITY: S-adenosyl-L-methionine + putrescine = S-  
 CC adenosyl-L-homocysteine + N-methylputrescine.  
 CC -1- PATHWAY: Nicotine biosynthesis.  
 CC -1- SIMILARITY: BELONGS TO THE SPERMIDINE/SPERMINE SYNTHASES FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL, AF126812; AAF14881.1; -;  
 DR InterPro; IPR000051; SAM\_bind.  
 DR InterPro; IPR001045; Spermine synthase.  
 DR Pfam; PF01564; Spermine synth; 1.  
 DR TIGRFAMs; TIGR00417; speE; 1.  
 DR PROSITE; PS01330; SPERMIDINE SYNTHASE; 1.  
 DR TRANSFAM; Methyltransferase; Multigene family.  
 SQ SEQUENCE 419 AA; 45899 MW; BDBE4417C454ED16 CRC64;  
 QY 95 ENGRFPY 101  
 DB 185 ENGRFPY 191

RESULT 17  
 ID MSRA\_HAEIN STANDARD; PRT; 445 AA.  
 AC P5164;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Protein mra homolog.  
 GN (MSRA-A OR H11337) AND (MSRA-B OR H11463).  
 OS Haemophilus influenzae.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;  
 OC Pasteurellaceae; Haemophilus.  
 NCBI\_TaxID=727;  
 OK NCBI\_TaxID=727;  
 RN  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=RD / KW20 / ATCC 51907;  
 RX MEDLINE=95350630; PubMed=7542800;  
 RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kitzner E.F.,  
 RA Kierleberg A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick E.F.,  
 RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,  
 RA Scott J.D., Shrivley R., Liu L.-I., Glodok A., Kelley J.M.,  
 RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,  
 RA Uterback T.R., Hanna M.C., Nguyen D.T., Sauder D.M., Brandon R.C.,  
 RA Fine L.D., Fritchman J.L., Puhmann J.L., Geoghagen N.S.M.,  
 RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,  
 RA "Whole-genome random sequencing and assembly of Haemophilus influenzae  
 Rd.";  
 RT Science 269:496-512(1995).  
 RL  
 CC -1- SIMILARITY: BELONGS TO THE PHOSPHOHEXOSE MUTASES FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC -----  
 CC EMBL; U32813; AAC22984.1; -  
 CC EMBL; U32824; AAC23110.1; -  
 CC PIR; C64172; C64172.  
 CC  
 CC TIGR; H11337; -  
 CC  
 CC DR InterPro; IPR006352; GIMM.  
 CC DR InterPro; IPR005841; PG\_PMM\_mutase.  
 CC DR InterPro; IPR005844; PG\_PMM\_ABAI.  
 CC DR InterPro; IPR005845; PG\_PMM\_ABAI.  
 CC DR InterPro; IPR005846; PG\_PMM\_ABAI.  
 CC DR InterPro; IPR005843; PG\_PMM\_C.  
 CC DR Pfam; PF00408; PGM\_PMM\_1.  
 CC DR Pfam; PF02878; PGM\_PMM\_1; 1.  
 CC DR Pfam; PF02879; PGM\_PMM\_1; 1.  
 CC DR Pfam; PF02880; PGM\_PMM\_III; 1.  
 CC DR PRINTS; PR00509; PGM\_PMM.  
 CC DR TIGR; TIGR01455; GIMM; 1.  
 CC DR PROSITE; PS00710; PGM\_PMM; 1.  
 CC KW Phosphorylation; Complete proteome.  
 CC FT ACT\_SITE 102 102  
 CC PHOSPHOSERINE INTERMEDIATE  
 CC (BY SIMILARITY).  
 CC  
 CC SQ SEQUENCE 445 AA; 47372 MW; 0335916382PFDRI CRC64;

Query Match 5.3%; Score 7; DB 1; Length 445;  
 Best Local Similarity 100.0%; Pred. No. 25;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 57 ELASAVK 63  
 DB 361 ELASAVK 367  
 RESULT 18

SYN HALNT  
 ID SYN HALNT STANDARD; PRT; 494 AA.  
 AC 09H09;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE Cysteineyl-tRNA synthetase (EC 6.1.1.16) (Cysteine--tRNA ligase)  
 DE (CysRS)  
 GN CYS OR VNC1097G.  
 OS Halobacterium sp. (strain NRC-1 / ATCC 700922 / JCM 11081).  
 OC Archaea; Euryarchaeota; Halobacteria; Halobacteriales;  
 OC Halobacteriaceae; Halobacterium.  
 OC NCBI\_TaxID=64091;  
 OK NCBI\_TaxID=64091;  
 RN  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=20504483; PubMed=11016950;  
 RA Ng W.V., Kennedy S.P., Mahairas G.G., Bergquist B., Pan M.,  
 RA Shukla H.D., Lasky S.R., Baliga N.S., Thorsson V., Sbrogna J.,  
 RA Swartzell S., Weir D., Hall J., Dahl T.A., Welti R., Goo Y.A.,  
 RA Leitbauer B., Keller K., Cruz R., Danon M.J., Hough D.W.,  
 RA Maddock D.G., Jablonski P.E., Krebs M.P., Angevine C.M., Dale H.,  
 RA Isenberger T.A., Beck R.F., Pohlshroder M., Spudich J.L., Jung K.-H.,  
 RA Alam W., Freitas T., Hou S., Daniels C.J., Dennis P.P., Omer A.D.,  
 RA Ebhardt H., Lowe T.M., Liang P., Riley M., Hood L., Dassarma S.;  
 RA "Genome sequence of Halobacterium species NRC-1.";  
 RT Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181(2000).  
 RL  
 CC -1- CATALYTIC ACTIVITY: ATP + L-cysteine + tRNA(Cys) = AMP +  
 CC diphosphate + L-cysteineyl-tRNA(Cys).  
 CC  
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.  
 CC  
 CC -1- SIMILARITY: Belongs to class-I aminoacyl-tRNA synthetase family.  
 CC strong, to methionyl-tRNA synthetase.  
 CC  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC -----  
 CC EMBL; AB005041; AAC19494.1; -  
 CC PIR; B84266; B84266.  
 CC DR HAMAP; MF\_00041; -; 1.  
 CC DR InterPro; IPR002308; Cys\_tRNA-synt\_1a.  
 CC DR InterPro; IPR001412; tRNA-synt\_1.  
 CC DR Pfam; PF01406; tRNA-synt\_1a; 1.  
 CC DR PRINTS; PR00983; TRNASYNTHCS.  
 CC DR TIGR; TIGR00435; CYS; 1.  
 CC DR PROSITE; PS00178; AA\_TRNA\_LIGASE\_1; FALSE NEG.  
 CC KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;  
 CC KW Complete proteome.  
 CC FT SITE 31 41  
 CC "HIGH" REGION.  
 CC FT SITE 287 291  
 CC "KMSKS" REGION.  
 CC  
 CC SQ SEQUENCE 494 AA; 54034 MW; 2EBB9461A557D87 CRC64;

Query Match 5.3%; Score 7; DB 1; Length 494;  
 Best Local Similarity 100.0%; Pred. No. 27;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 56 LELASAV 62  
 DB 394 LELASAV 400  
 RESULT 19  
 COE2 BRARE STANDARD; PRT; 579 AA.  
 ID COE2 BRARE  
 AC 093375;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Transcription factor COE2.  
 GN COE2.

```

OS Brachydanio rerio (Zebrafish) (Danio rerio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Danio.
OC NCBI_TaxID=7955;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99002949; PubMed=9784615;
RA Bailey-Cul L., Dubois L., Vincent A.;
RT "Molecular cloning of Xcoez2, the zebrafish homolog of Xenopus Xcoez2
RT and mouse ESF-2, and its expression during primary neurogenesis.";
RL Mech. Dev. 77:85-90(1998).
CC -1- SUBCELLULAR LOCATION: Nuclear (potential).
CC -1- DEVELOPMENTAL STAGE: First detected at the 75% epiboly stage where
CC it covers the anterior neural plate. Widely expressed in the
CC presumptive mesencephalon and rhombomeres 1-4 until the 2-3-somite
CC stage, with expression persisting in ngn1-positive clusters. First
CC detected in the olfactory placodes at the 5-somite stage. In the
CC spinal cord, detected in ngn1-positive clusters of primary
CC neuroblasts during the early somite stages. Expression decreases
CC in the spinal cord from the 30-somite stage but persists in the
CC olfactory bulb and regions of the rhombencephalon and brain.
CC -1- SIMILARITY: BELONGS TO THE COE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF072657; AAC96103.1; -.
DR ZFIN; ZDB-GENE-990715-11; coez2.
DR InterPro; IPR001092; HLH_basisc.
DR InterPro; IPR002909; IPT_TIG.
DR InterPro; IPR003523; TF_COE.
DR Pfam; PF01833; TIG; 1.
DR SMART; SM00353; HLH; 1.
DR SMART; SM00429; IPT; 1.
DR PROSITE; PS01345; COE; 1.
KW Transcription regulation; DNA-binding; Nuclear protein;
KW Developmental protein; Zinc-finger.
FT ZN_FING 149 168
FT DOMAIN 460 537
FT SER/THR/PRO-RICH
SQ SEQUENCE 579 AA; 63528 MW; E2DD1919055EEA43 CRC64;

```

```

Query Match 5.3%; Score 7; DB 1; Length 579;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 19 GEGGOTS 25
Db 446 GEGGOTS 452

```

```

RESULT 20
YFES SCHPO STANDARD; PRT; 601 AA.
AC P87132;
DT 15-JUL-1998 (Rel. 36, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hypothetical protein C167.05 in chromosome I.
GN SPAC167.05 OR SPAC57A7.01..
OS Schizosaccharomyces pombe (fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972;

```

```

RX MEDLINE=21848401; PubMed=11859360;
RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA Sounos J., Peat N., Hayles J., Baker S., Baaham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Felwell T., Frazer A.,
RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagals K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Ocell C.,
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabbittowitch B.,
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodard J., Wolckart G., Aert R., Robben J., Grymonprez B.,
RA Welfens I., Vanreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,
RA Borzym K., Langer I., Beck A., Lehnach H., Reinhardt R., Pohl T.M.,
RA Egger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA Goffeau A., Cadieu E., Dreno S., Gloux S., Lelaure V., Mottier S.,
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Rochet M., Galliard C., Tallada V.A., Garzon A., Rhode G.,
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA Shpakovski G.V., Ussery D., Barrett B.G., Nurse P.;
RA "The genome sequence of Schizosaccharomyces pombe.";
RL Nature 415:871-880(2002).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AL035248; CAA22850.1; -.
DR EMBL; Z95396; CAB08759.1; -.
DR PIR; T37738; T37738.
DR GenDB_Spombe; SPAC167.05; -.
DR InterPro; IPR006015; USP.
DR InterPro; IPR006016; USP_dom.
DR Pfam; PF00582; USP; 1.
DR PRINTS; PR01438; UNWRSLSTRESS.
KW Hypothetical protein.
SQ SEQUENCE 601 AA; 65947 MW; 13FAACBDC14239 CRC64;

```

```

Query Match 5.3%; Score 7; DB 1; Length 601;
Best Local Similarity 100.0%; Pred. No. 32;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 116 ETELEKIT 122
Db 487 ETELEKIT 493

```

```

RESULT 21
PBEP_BACSU STANDARD; PRT; 716 AA.
AC Q07668;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Penicillin-binding protein 2B (PBp-2B).
GN PBEP.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=94064553; PubMed=8244929;
RA Yanouri A., Daniel R.A., Eyrington J., Buchanan C.E.;

```

```

RT      "Cloning and sequencing of the cell division gene pbpb, which encodes
RT      penicillin-binding protein 2B in Bacillus subtilis."
RT      J. Bacteriol. 175:7604-7616(1993).
RN      [2]
RP      SEQUENCE FROM N.A.
RC      STRAIN=168;
RC      Daniel R.A., Williams A.M., Errington J.;
RL      Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.
RN      [3]
RP      SEQUENCE FROM N.A.
RC      STRAIN=168;
RC      MEDLINE=98044033; PubMed=9384377;
RC      Kuntz F., Ogasawara N., Mosser I., Albertini A.M., Alloni G.,
RC      Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Bocheret S.,
RC      Borries R., Bourestel L., Brans A., Braun M., Brignelli S.C., Bron S.,
RC      Brouillet S., Brusch C.V., Caldwell B., Capuano V., Carter N.M.,
RC      Choi S.K., Codani J.J., Conerton I.F., Cummings N.J., Daniel R.A.,
RC      Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,
RC      Eutlian K.D., Errington J., Fabre C., Ferrati E., Foulger D.,
RC      Fritz C., Fujita M., Fujita Y., Fuma S., Gallizzi A., Galleron N.,
RC      Ghim S.Y., Glaeser P., Goffeau A., Golligly E.J., Grandi G.,
RC      Giuseppe G., Guy B.J., Haga K., Hatach J., Harwood C.R., Henaut A.,
RC      Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RC      Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RC      Kobayashi Y., Koeltter P., Koningsstein G., Krogh S., Kumano M.,
RC      Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RC      Lee S.M., Levine A., Liu H., Maunda S., Maue C., Medigue C.,
RC      Medina N., Mellado R.P., Mizuno M., Mostl D., Nakai S., Noback M.,
RC      Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RC      Parro V., Pohl T.M., Portetelle D., Portwolk S., Prescott A.M.,
RC      Priebe E., Puig P., Furnelle B., Rapoport G., Rey M., Reynolds S.,
RC      Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y.,
RC      Sato T., Scanlan E., Schleich S., Schroeter R., Scottone F.,
RC      Sekiguchi J., Sekowska A., Seror S.J., Serro P., Shin B.S., Soldo B.,
RC      Sorochin A., Taccini E., Takagi T., Takahashi H., Takemura K.,
RC      Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Toironi A.,
RC      Tosato V., Uchiyama S., Vandenbol M., Vanier F., Vassarotti A.,
RC      Viari A., Wambutt R., Wedler H., Wedler H., Weitzneger T.,
RC      Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RC      Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa K., Danchin A.;
RT      "The complete genome sequence of the Gram-positive bacterium Bacillus
RT      subtilis".
RT      Nature 390:249-256(1997).
RN      [4]
RP      SEQUENCE OF 675-716 FROM N.A.
RC      STRAIN=168;
RC      MEDLINE=94118264; PubMed=8289242;
RC      Daniel R.A., Drake S., Buchanan C.E., Scholle R., Errington J.;
RT      "The Bacillus subtilis spvD gene encodes a mother-cell-specific
RT      penicillin-binding protein required for spore morphogenesis."
RT      J. Mol. Biol. 235:209-220(1994).
CC      -1- FUNCTION: PENICILLIN-BINDING PROTEINS (PBPS) FUNCTION IN THE
CC      LATE STEPS OF MUREIN BIOSYNTHESIS. PBp-2B IS REQUIRED FOR
CC      VEGETATIVE CELL DIVISION AND SPOULATION SEPARATION. BETA-
CC      LACTAMASE INACTIVATES THE PBPS BY ACTYLATING AN ESSENTIAL SERINE
CC      RESIDUE IN THE ACTIVE SITE OF THESE PROTEINS, THEREBY INTERRUPTING
CC      NORMAL CELL WALL SYNTHESIS.
CC      -1- SUBCELLULAR LOCATION: Type II membrane protein.
CC      -1- DEVELOPMENTAL STAGE: SYNTHESIZED THROUGHOUT VEGETATIVE GROWTH.
CC      SYNTHESIS IS ENHANCED DURING STAGE II OF SPOULATION.
CC      UNDETECTABLE IN THE MATURE SPORE.
CC      -1- SIMILARITY: BELONGS TO THE TRANSPEPTIDASE FAMILY.
CC      -1- CAUTION: IT IS UNCERTAIN WHETHER MET-1 OR MET-4 IS THE INITIATOR.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
CC      EMBL; L09703; AAC36837.1; -

```

```

DR      EMBL; Z68230; CAA2527.1; -
DR      EMBL; Z99111; CAB1389.1; -
DR      EMBL; Z25865; CAA81084.1; -
DR      PIR; C53292; C53292.
DR      Subtilisin; BG10221; pbpb.
DR      InterPro; IPR005543; PASTA.
DR      InterPro; IPR005311; pbp dimer.
DR      InterPro; IPR001460; Transpeptidase.
DR      Pfam; PF03793; PASTA; 2.
DR      Pfam; PF03717; pbp_dimer; 1.
DR      Pfam; PF00905; Transpeptidase; 1.
DR      SMART; SM00740; PASTA; 2.
KW      Peptidoglycan synthetase; Cell division; Antibiotic resistance;
KW      Cell wall; Transmembrane; Sporulation; Complete proteome.
KW      CYTOPLASMIC (POTENTIAL).
FT      DOMAIN 1 13
FT      TRANSMEM 14 31
FT      DOMAIN 32 312
FT      ACT_SITE 312 312
FT      ACT_SITE 716 716
SQ      SEQUENCE 716 AA; 79305 MW; 929B1BD591D50B4 CRC64;
Query Match 5.3%; Score 7; DB 1; Length 716;
Best Local Similarity 100.0%; Pred.No. 38;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY      61 AVKEQYP 67
DB      633 AVKEQYP 639
RESULT 22
MUTS_CHLTR STANDARD; PRI; 820 AA.
AC 084757;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE DNA mismatch repair protein muts.
DN DNA mismatch repair protein muts.
GN MUTS OR CT792.
OS Chlamydia trachomatis.
OC Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=813;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=D/UM-3/CX;
RC MEDLINE=99000809; PubMed=9784136;
RC Stephens R.S., Kaiman S., Tammel C.J., Fan J., Marathe R., Aravind L.,
RC Mitchell W.P., Ollinger L., Tatunov R.L., Zhao Q., Koonin E.V.,
RA Davis R.W.;
RT "Genome sequence of an obligate intracellular pathogen of humans:
RT Chlamydia trachomatis."
RL Science 282:754-759(1998).
CC -1- FUNCTION: This protein is involved in the repair of mismatches in
CC DNA. It is possible that it carries out the mismatch recognition
CC step. This protein has a weak ATPase activity (by similarity).
CC -1- SIMILARITY: BELONGS TO THE DNA MISMATCH REPAIR MUTS FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AE001351; AAC68387.1; -
DR PIR; D71471; D71471.
DR HAMAP; MF_00096; -; 1.
DR InterPro; IPR005748; MutS1.
DR InterPro; IPR000432; MutS_C.
DR InterPro; IPR002863; MutS_N.
DR Pfam; PF01624; MutS_1; 1.
DR Pfam; PF05188; MutS_1; 1.
DR Pfam; PF05192; MutS_1; 1.

```

DR Pfam; PF05190; Muts\_IV; 1.  
 DR Pfam; PF00488; Muts\_V; 1.  
 DR ProDom; PD001263; Muts\_C; 1.  
 DR SMART; SM00534; Mutsac; 1.  
 DR SMART; SM00533; Mutsd; 1.  
 DR TIGRfam; TIGR01070; muts1; 1.  
 DR PROSITE; PS00486; DNA\_MISMATCH\_REPAIR\_2; 1.  
 DR DNA repair; ATP-binding; DNA-binding; Complete proteome.  
 NP BIND 618 625 ATP (POTENTIAL).  
 SQ SEQUENCE 820 AA; 92130 MM; A5EF7B2BED3A85A5 CRC64;

Query Match 5.3%; Score 7; DB 1; Length 820;  
 Best Local Similarity 100.0%; Pred. No. 42;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 103 KDLIEAI 109  
 Db 154 KDLIEAI 160

RESULT 23  
 POLG LIV STANDARD; PRT; 980 AA.  
 ID POLG LIV  
 AC P23338;  
 DT 01-AUG-1991 (Rel. 19, Created)  
 DT 01-AUG-1991 (Rel. 19, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Genome polypeptide [Contains: Capsid protein C (core protein); Matrix protein (Envelope protein M); Major envelope protein E; Nonstructural proteins NS1] (Fragment).  
 DE Louping ill virus (LI).  
 OS Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
 OC Flavivirus.  
 NC NCB1\_Taxid=11086;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=91082437; PubMed=1845834;  
 RA Shiu S.Y.W., Ayres M.D., Gould E.A.;  
 RT "Genomic sequence of the structural proteins of louping ill virus: comparative analysis with tick-borne encephalitis virus.";  
 RL Virology 180:411-415(1991).  
 CC -I- SUBUNIT. THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A LIPID PROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF PROTEIN C AND MRNA.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.1sb-sib.ch/announce/> or send an email to [license@1sb-sib.ch](mailto:license@1sb-sib.ch)).  
 CC EMBL; MS9376; AAA46281.1; -  
 DR PIR; A38523; A38523.  
 DR HSSP; P14336; 1SVB.  
 DR InterPro; IPR001122; Flavi\_CapsidC.  
 DR InterPro; IPR000336; Flavi\_glycoprote.  
 DR InterPro; IPR000069; Flavi\_M.  
 DR InterPro; IPR001157; Flavi\_NS1.  
 DR InterPro; IPR002535; Flavi\_propep.  
 DR Pfam; PF01003; Flavi\_capsid; 1.  
 DR Pfam; PF02832; Flavi\_glycop\_C; 1.  
 DR Pfam; PF00869; Flavi\_glycoprot; 1.  
 DR Pfam; PF01004; Flavi\_M; 1.  
 DR Pfam; PF00948; Flavi\_NS1; 1.  
 DR Pfam; PF01570; Flavi\_propep; 1.  
 DR ProDom; PD001556; Flavi\_glycoprote; 1.  
 DR ProDom; PD001496; Flavi\_NS1; 1.  
 DR Polyprotein; Glycoprotein; Core protein; Coat protein; Envelope protein; Transmembrane; Nonstructural protein.  
 KW INIT\_MET 1 1 REMOVED FROM CAPSID PROTEIN C BY THE

FT FT CHAIN 1 112  
 FT PROPEP 113 205  
 FT CHAIN 113 205  
 FT CHAIN 206 280  
 FT CHAIN 281 776  
 FT CHAIN 777 980  
 FT TRANSMEM 101 112  
 FT TRANSMEM 247 259  
 FT TRANSMEM 266 280  
 FT TRANSMEM 738 751  
 FT DISULFID 283 310  
 FT DISULFID 340 396  
 FT DISULFID 354 385  
 FT DISULFID 372 401  
 FT DISULFID 466 570  
 FT DISULFID 587 618  
 FT CARBOHYD 144 144  
 FT CARBOHYD 434 434  
 FT CARBOHYD 861 861  
 FT NON TER 980 980  
 SQ SEQUENCE 980 AA; 107234 MM; 2242B9364F3BFD9 CRC64;

Query Match 5.3%; Score 7; DB 1; Length 980;  
 Best Local Similarity 100.0%; Pred. No. 49;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 58 LASAYKE 64  
 Db 818 LASAYKE 824

RESULT 24  
 XPG MOUSE STANDARD; PRT; 1170 AA.  
 ID XPG MOUSE  
 AC P35689; Q61528; Q64248;  
 DT 01-JUN-1994 (Rel. 29, Created)  
 DT 15-JUN-1999 (Rel. 38, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE DNA-repair protein complementing XP-G cells homolog (Xeroderma pigmentosum group G complementing protein homolog) (DNA excision repair protein ERCC-5).  
 DE ERCC5 OR XPG OR ERCC-5.  
 GN ERCC5 OR XPG OR ERCC-5.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 NC NCB1\_Taxid=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=94173288; PubMed=7510366;  
 RA Shiom T., Harada Y.-N., Saito T., Shiom N., Okuno Y., Yamazumi M.;  
 RT "An ERCC5 gene with homology to yeast Rad2 is involved in group G Xeroderma pigmentosum.";  
 RL Mutat. Res. 314:167-175(1994).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/10; TISSUE=Liver;  
 RX MEDLINE=96070433; PubMed=7590748;  
 RA Harada Y.N., Matsuda Y., Shiom N., Shiom T.;  
 RT "Complementary DNA sequence and chromosomal localization of xpg, the mouse counterpart of human repair gene XPG/ERCC5.";  
 RL Genomics 28:59-65(1995).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=DBA/2;  
 RX MEDLINE=96559149; PubMed=8703115;  
 RA Ludwig D.L., Mudgett J.S., Park M.S., Perez-Castro A.V., MacInnes M.A.;  
 RT "Molecular cloning and structural analysis of the functional mouse genomic Xpg gene.";  
 RL Mamm. Genome 7:644-649(1996).  
 CC -I- FUNCTION: SINGLE-STRANDED DNA ENDONUCLEASE INVOLVED IN DNA EXCISION REPAIR. MAKES THE 3' INCISION IN DNA NUCLEOTIDE EXCISION REPAIR (BY SIMILARITY).  
 CC

CC -1- SUBCELLULAR LOCATION: Nuclear (Probable).  
 CC -1- SIMILARITY: BELONGS TO THE XPG/RAD2 ENDONUCLEASE FAMILY. XPG  
 CC SUBFAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: D16306; BAA03813.1; -;  
 DR EMBL: U40796; AAA91039.1; -;  
 DR EMBL: U40795; AAB17885.1; -;  
 DR EMBL: U39892; AAB17885.1; JOINED.  
 DR EMBL: U39893; AAB17885.1; JOINED.  
 DR EMBL: U39894; AAB17885.1; JOINED.  
 DR EMBL: U39896; AAB17885.1; JOINED.  
 DR EMBL: U40073; AAB17885.1; JOINED.  
 DR EMBL: U40431; AAB17885.1; JOINED.  
 DR EMBL: U40432; AAB17885.1; JOINED.  
 DR EMBL: U40668; AAB17885.1; JOINED.  
 DR EMBL: U40670; AAB17885.1; JOINED.  
 DR EMBL: U40792; AAB17885.1; JOINED.  
 DR EMBL: U40793; AAB17885.1; JOINED.  
 DR EMBL: U40794; AAB17885.1; JOINED.  
 DR PIR: A57650; A57650.  
 DR MGD: MGI:103582; Erc5.  
 DR InterPro: IPR000513; E50\_N1.  
 DR InterPro: IPR003584; HH2\_2.  
 DR InterPro: IPR006086; XPG\_I.  
 DR InterPro: IPR006085; XPG\_N.  
 DR InterPro: IPR001044; XPGC\_DNA\_repair.  
 DR InterPro: IPR006084; XPGC\_Rad.  
 DR Pfam: PF00867; XPG\_I; 1.  
 DR Pfam: PF00752; XPG\_N; 1.  
 DR PRINTS: PR00853; XPGRADSUPER.  
 DR SMART: SM00279; HH2; 1.  
 DR SMART: SM00484; XPGI; 1.  
 DR SMART: SM00485; XPGN; 1.  
 DR TIGRFAMs: TIGR00600; rad2; 1.  
 DR PROSITE: PS00841; XPG\_1; 1.  
 DR PROSITE: PS00842; XPG\_2; 1.  
 DR DNA\_repair; DNA-binding; Nuclear protein; Hydrolyase; Nuclease;  
 KW Endonuclease.  
 FT DOMAIN 1 95 N-DOMAIN.  
 FT DOMAIN 752 882 I-DOMAIN.  
 FT DOMAIN 1049 1065 NUCLEIC ACID LOCALIZATION SIGNAL (POTENTIAL).  
 FT VARIANT 388 388 R -> C (IN STRAIN DBA/2).  
 FT VARIANT 488 488 S -> R (IN STRAIN DBA/2).  
 FT VARIANT 688 688 I -> T (IN STRAIN DBA/2).  
 FT VARIANT 1015 1015 S -> N (IN STRAIN DBA/2).  
 FT VARIANT 1021 1021 M -> I (IN STRAIN DBA/2).  
 FT VARIANT 1121 1121 S -> P (IN STRAIN DBA/2).  
 FT CONFLICT 227 227 N -> M (IN REF. 1).  
 FT CONFLICT 249 249 N -> M (IN REF. 1).  
 FT CONFLICT 300 302 VMD -> MDE (IN REF. 1).  
 FT CONFLICT 313 313 N -> M (IN REF. 1).  
 FT CONFLICT 320 320 N -> M (IN REF. 1).  
 FT CONFLICT 339 339 N -> M (IN REF. 1).  
 FT CONFLICT 408 408 D -> DVOTGG (IN REF. 1).  
 FT CONFLICT 581 581 N -> NSASEVIGEV (IN REF. 1).  
 FT CONFLICT 795 795 M -> V (IN REF. 1).  
 FT CONFLICT 1039 1039 A -> AMEKKEFL (IN REF. 1).  
 FT CONFLICT 1134 1134 S -> SD (IN REF. 1).  
 FT CONFLICT 1157 1158 KL -> RR (IN REF. 1).  
 SO SEQUENCE 1170 AA, 130864 MW, 4058670FPD1770ED CRC64;

Query Match 5.3%; Score 7; DB 1; Length 1170;  
 Best Local Similarity 100.0%; Pred. No. 58;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 58 LASAYKE 64  
 DB 486 LASAYKE 492  
 RESULT 25  
 DME\_ARATH ID DME\_ARATH STANDARD; PRT; 1729 AA.  
 AC 08L56; 09L267; 09L268; 09L269;  
 DT 15-SEP-2003 (Rel. 42, Created)  
 DT 15-SEP-2003 (Rel. 42, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE Transcriptional activator DEMETER (DNA glycosylase-related protein  
 DME).  
 GN DME OR AT5G04560/AT5G04570/AT5G04580 OR  
 GN T12M21.160/T32M21.170/T32M21.180.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
 OX NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A. (ISOFORM 1), FUNCTION, AND CHARACTERIZATION.  
 RC STRAIN=cv. Columbia; TISSUE=flower;  
 RX MEDLINE=22145911; PubMed=12150995;  
 RA Choi Y., Gehring M., Johnson L., Hannon M., Harada J.U.,  
 RA Goldberg R.B., Jacobsen S.E., Fischer R.L.;  
 RT "DEMETER, a DNA glycosylase domain protein, is required for endosperm  
 RT gene imprinting and seed viability in Arabidopsis.";  
 RL Cell 110:33-42(2002).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. Columbia;  
 RX MEDLINE=21016721; PubMed=1130714;  
 RA Tabata S., Kaneko T., Nakamura Y., Kotani H., Kato T., Asamizu E.,  
 RA Miyajima N., Sasamoto S., Kimura T., Hosouchi T., Kawashima K.,  
 RA Kohara M., Matsumoto M., Matsuno A., Muraki A., Nakayama S.,  
 RA Nakazaki N., Naito K., Okumura S., Shinpo S., Takeuchi C., Wada T.,  
 RA Matanabe A., Yamada M., Yasuda M., Sato S., de la Bastide M.,  
 RA Huang E., Spiegel L., Gnoj L., O'Shaughnessy A., Preston R.,  
 RA Habermann K., Murray J., Johnson D., Rohlfing T., Nelson R.,  
 RA Stoneking T., Pepin K., Spieth J., Sekhon M., Armstrong J., Becker M.,  
 RA Belter E., Cordum H., Cordes M., Courtney L., Courtney J., Dante M.,  
 RA Du H., Edwards J., Fryman J., Haakensen B., Lamar E., Latreille P.,  
 RA Leonard S., Meyer R., Mulvaney E., Ozersky P., Riley A., Strommatt C.,  
 RA Wagner-McPherson C., Wollam A., Yeakum M., Bell M., Dedhia N.,  
 RA Parnell L., Shah R., Rodriguez M., Hoon See L., Vil D., Baker J.,  
 RA Kirchhoff K., Toth K., King L., Bahret A., Miller B., Marra M.,  
 RA Martienssen R., McCombie W.R., Wilson R.K., Murphy G., Bancroft I.,  
 RA Volckaert G., Wambutt R., Duesterhoeft A., Stiekema W., Pohl T.,  
 RA Ertlan K.-D., Terry N., Hartley N., Bent E., Johnson S.,  
 RA Langham S.-A., McCullagh B., Robben J., Grymonprez B., Zimmermann W.,  
 RA Ramsperger U., Wedler E., Balke K., Wedler E., Peters S.,  
 RA van Staveren T., Dirke W., Moolman P., Klein Lankhorst R.,  
 RA Weitzenecker T., Bothe G., Rose M., Hauf J., Berleider S., Hempel S.,  
 RA Feldpausch M., Lamberth S., Villarroel R., Gielens J., Ardiles W.,  
 RA Bents O., Lemcke K., Kolesov G., Mayer K.F.X., Rudd S., Schoof H.,  
 RA Schueller C., Zaccaria P., Newes H.-W., Bevan M., Franz P.F.;  
 RT "Sequence and analysis of chromosome 5 of the plant Arabidopsis  
 thaliana.";  
 RL Nature 408:823-826(2000).  
 RN [3]  
 RP SEQUENCE FROM N.A. (ISOFORM 2).  
 RC STRAIN=cv. Columbia;  
 RA Seki M., Iida K., Satou M., Sakurai T., Akiyama K., Ishida J.,  
 RA Nakajima M., Enju A., Kamiya A., Narusaka M., Carninci P., Kawai J.,  
 RA Hayashizaki Y., Shinozaki K.;  
 RT "Arabidopsis thaliana full-length cDNA.";  
 RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: Transcriptional activator involved in gene imprinting.  
 CC Allows the expression of the maternal copy of the imprinted MEA  
 CC gene before fertilization, possibly by antagonizing or suppressing





BELONG TO THE "LOW-VOLTAGE ACTIVATED (LVA)" GROUP AND ARE STRONGLY BLOCKED BY NICKEL AND MIBERRADIL. A PARTICULARITY OF THIS TYPE OF CHANNELS IS AN OPENING AT QUITE NEGATIVE POTENTIALS, AND A VOLTAGE-DEPENDENT INACTIVATION. T-TYPE CHANNELS SERVE PACEMAKING FUNCTIONS IN BOTH CENTRAL NEURONS AND CARDIAC NODAL CELLS AND SUPPORT CALCIUM SIGNALING IN SECRETORY CELLS AND VASCULAR SMOOTH MUSCLE. THEY MAY ALSO BE INVOLVED IN THE MODULATION OF FIRING PATTERNS OF NEURONS WHICH IS IMPORTANT FOR INFORMATION PROCESSING AS WELL AS IN CELL GROWTH PROCESSES.

-1- SUBCELLULAR LOCATION: Integral membrane protein.

-1- ALTERNATIVE PRODUCTS:

Name=1; Synonyms=Alh-a; IsoId=095180-1; Sequence=Displayed;

Name=2; Synonyms=Alh-b; IsoId=095180-2; Sequence=VSP\_000949;

-1- TISSUE SPECIFICITY: Expressed in kidney, liver, heart, brain. Isoform 2 seems to be testis-specific.

-1- DOMAIN: EACH OF THE FOUR INTERVAL REPEATS CONTAINS FIVE HYDROPHOBIC TRANSMEMBRANE SEGMENTS (S1, S2, S3, S5, S6) AND ONE POSITIVELY CHARGED TRANSMEMBRANE SEGMENT (S4). S4 SEGMENTS PROBABLY REPRESENT THE VOLTAGE-SENSOR AND ARE CHARACTERIZED BY A SERIES OF POSITIVELY CHARGED AMINO ACIDS AT EVERY THIRD POSITION.

-1- PTM: IN RESPONSE TO RAISING OF INTRACELLULAR CALCIUM, THE T-TYPE CHANNELS ARE ACTIVATED BY CAM-KINASE II.

-1- SIMILARITY: BELONGS TO THE CALCIUM CHANNEL ALPHA-1 SUBUNITS FAMILY.

-1- CAUTION: AT-AC pre mRNA splicing gives rise to the isoform 1 shown in this entry. The additional 20 amino acids found in the Ref.4 and Ref.6 sequences are due to a misunderstanding of the real type of splicing mechanism involved.

-----

THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

-----

CC EMBL: AF051946; AAC67239.3; -

CC EMBL: AF073931; AAD17668.1; -

CC EMBL: AEO06466; AAK61268.1; ALT\_SEQ.

CC EMBL: A0420779; CAD12646.1; -

CC EMBL: AL031703; CAC42094.1; ALT\_SEQ.

CC EMBL: AF223562; AAF60162.1; -

CC EMBL: AF223563; AAF60163.1; -

CC Genew: HGNC.1395; CACNA1H.

DR GK: 095180.1;

DR GO: GO:0005891; C:voltage-gated calcium channel complex; TAS.

DR GO: GO:0008332; F:low voltage-gated calcium channel activity; TAS.

DR GO: GO:0006936; P:muscle contraction; TAS.

DR GO: GO:0007520; P:myoblast fusion; TAS.

DR GO: GO:0008016; P:regulation of heart; TAS.

DR GO: GO:0006832; P:small molecule transport; TAS.

DR InterPro: IPR001682; Ca/Na pore.

DR InterPro: IPR002111; Cat\_channel\_TripL.

DR InterPro: IPR005821; Ion trans.

DR InterPro: IPR005820; M-channel nlg.

DR InterPro: IPR005445; TVDCCA1phal.

DR Pfam: PF00520; Ion\_trans\_4.

DR PRINTS: PR01629; TVDCCA1phal.

KW Ionic channel; Transmembrane; Ion transport; Voltage-gated channel; Calcium channel; Glycoprotein; Repeat; Multigene family; Calcium-binding; Phosphorylation; Alternative splicing.

FT REPEAT 87 422 I

FT REPEAT 779 1018 II.

FT REPEAT 1281 1558 III.

FT REPEAT 1602 1863 IV.

FT DOMAIN 1 100 CYTOPLASMIC (POTENTIAL).

FT TRANSMEM 101 119 S1 OF REPEAT I (POTENTIAL).

FT DOMAIN 120 139 EXTRACELLULAR (POTENTIAL).

FT TRANSMEM 140 160 S2 OF REPEAT I (POTENTIAL).

FT	DOMAIN	161	169	CYTOPLASMIC (POTENTIAL).
FT	TRANSMEM	170	184	S3 OF REPEAT I (POTENTIAL).
FT	DOMAIN	185	193	EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	194	212	S4 OF REPEAT I (POTENTIAL).
FT	DOMAIN	213	232	CYTOPLASMIC (POTENTIAL).
FT	TRANSMEM	233	253	S5 OF REPEAT I (POTENTIAL).
FT	DOMAIN	254	394	EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	395	419	S6 OF REPEAT I (POTENTIAL).
FT	DOMAIN	420	793	CYTOPLASMIC (POTENTIAL).
FT	TRANSMEM	794	814	S1 OF REPEAT II (POTENTIAL).
FT	DOMAIN	815	827	EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	828	849	S2 OF REPEAT II (POTENTIAL).
FT	DOMAIN	850	855	CYTOPLASMIC (POTENTIAL).
FT	TRANSMEM	856	874	S3 OF REPEAT II (POTENTIAL).
FT	DOMAIN	875	882	EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	883	906	S4 OF REPEAT II (POTENTIAL).
FT	DOMAIN	907	917	CYTOPLASMIC (POTENTIAL).
FT	TRANSMEM	918	938	S5 OF REPEAT II (POTENTIAL).
FT	DOMAIN	939	990	EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	991	1015	S6 OF REPEAT II (POTENTIAL).
FT	DOMAIN	1016	1290	CYTOPLASMIC (POTENTIAL).
FT	TRANSMEM	1291	1313	S1 OF REPEAT III (POTENTIAL).
FT	DOMAIN	1314	1331	EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	1332	1352	S2 OF REPEAT III (POTENTIAL).
FT	DOMAIN	1353	1362	CYTOPLASMIC (POTENTIAL).
FT	TRANSMEM	1363	1382	S3 OF REPEAT III (POTENTIAL).
FT	DOMAIN	1383	1396	EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	1397	1418	S4 OF REPEAT III (POTENTIAL).
FT	DOMAIN	1419	1428	CYTOPLASMIC (POTENTIAL).
FT	TRANSMEM	1429	1452	S5 OF REPEAT III (POTENTIAL).
FT	DOMAIN	1453	1529	EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	1530	1555	S6 OF REPEAT III (POTENTIAL).
FT	DOMAIN	1556	1616	CYTOPLASMIC (POTENTIAL).
FT	TRANSMEM	1617	1637	S1 OF REPEAT IV (POTENTIAL).
FT	DOMAIN	1638	1651	EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	1652	1673	S2 OF REPEAT IV (POTENTIAL).
FT	DOMAIN	1674	1680	CYTOPLASMIC (POTENTIAL).
FT	TRANSMEM	1681	1699	S3 OF REPEAT IV (POTENTIAL).
FT	DOMAIN	1700	1713	EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	1714	1737	S4 OF REPEAT IV (POTENTIAL).
FT	DOMAIN	1738	1751	CYTOPLASMIC (POTENTIAL).
FT	TRANSMEM	1752	1772	S5 OF REPEAT IV (POTENTIAL).
FT	DOMAIN	1773	1835	EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	1836	1863	S6 OF REPEAT IV (POTENTIAL).
FT	DOMAIN	1864	2353	CYTOPLASMIC (POTENTIAL).
FT	TRANSMEM	2354	530	POLY-HIS.
FT	DOMAIN	531	530	POLY-SER.
FT	TRANSMEM	531	1110	CALCIUM ION SELECTIVITY AND PERMEABILITY (BY SIMILARITY).
FT	DOMAIN	1107	1110	CALCIUM ION SELECTIVITY AND PERMEABILITY (BY SIMILARITY).
FT	TRANSMEM	1583	1586	CALCIUM ION SELECTIVITY AND PERMEABILITY (BY SIMILARITY).
FT	DOMAIN	378	378	CALCIUM ION SELECTIVITY AND PERMEABILITY (BY SIMILARITY).
FT	TRANSMEM	974	974	CALCIUM ION SELECTIVITY AND PERMEABILITY (BY SIMILARITY).
FT	DOMAIN	1504	1504	CALCIUM ION SELECTIVITY AND PERMEABILITY (BY SIMILARITY).
FT	TRANSMEM	1808	1808	CALCIUM ION SELECTIVITY AND PERMEABILITY (BY SIMILARITY).
FT	DOMAIN	192	192	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	TRANSMEM	271	271	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	DOMAIN	1466	1466	N-LINKED (GLCNAC. . .) (POTENTIAL).

Query Match 5.3%; Score 7; DB 1; Length 2353;

Best Local Similarity 100.0%; Pred. No. 1; le-02;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 GPEAEP 15

Db 2172 GPEAEP 2178

RESULT 27

YH33 ARCFU STANDARD; PRT; 45 AA.

ID YH33 ARCFU

AC 028511;



```

DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein AF1733.
GN AF1733.
OS Archaeoglobus fulgidus.
OC Archaea; Euryarchaeota; Archaeoglobi; Archaeoglobales;
OC Archaeoglobaceae; Archaeoglobus.
OX NCBI_TaxID=2234;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VC-16 / DSM 4304 / ATCC 49558;
RX MEDLINE=98049343; PubMed=9389475;
RA Klein H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,
RA Ketchum K.A., Dodson R.J., Gwinn M., Hickey E.K., Peterson J.D.,
RA Richardson D.L., Kierlavage A.R., Graham D.E., Kyprides N.C.,
RA Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,
RA Kirkness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,
RA Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glodex A., Zhou L.,
RA Cotton M.D., Spriggs T., Artlich P., Kaine B.P., Sykes S.M.,
RA Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,
RA Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,
RA Venter J.C.;
RT "The complete genome sequence of the hyperthermophilic, sulphate-
RT reducing archaeon Archaeoglobus fulgidus.";
RL Nature 390:364-370(1997).
CC -1- SIMILARITY: STRONG, TO M.JANNASCHII MJ151 AND
CC M.THERMOAUTOTROPHICUM MTH1128.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AEO00983; AAB89518.1; --
CC DR PIR; D69466; D69466.
CC DR TIGR; AF1733; --
CC KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 45 AA; 5241 MW; 1BF65B53C9DD4B6A CRC64;
SQ
Query Match 4.6%; Score 6; DB 1; Length 45;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 39 VRIVE 44
Db 5 VRIVE 10

```

```

RN [2]
RP SEQUENCE FROM N.A.
RA Salvatore C.A., Jacobson M.A.;
RL Submitted (XXX-1992) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=99222459; PubMed=10198197;
RA McTierman C.F., Frye C.S., Lemster B.H., Kinder E.A.,
RA Ogletree-Hughes M.L., Moravec C.S., Feldman A.M.;
RT "The human phospholamban gene: structure and expression.";
RL J. Mol. Cell. Cardiol. 31:679-692(1999).
RN [4]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=92388257; PubMed=1247932;
RA Struhsberg R.L., Feigold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shennan C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buettow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.U., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hultky S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Ketterman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butcherfield Y.S.N., Krzywinski M.I., Skalska U., Smalhus D.E.,
RA Schnerck A., Schein J.E., Jones S.J.M., Marx M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [5]
RP STRUCTURE BY NMR OF 1-25.
RX MEDLINE=95298769; PubMed=7779806;
RA Mortishire-Smith R.J., Pitzemberger S.M., Burke C.J., Middaugh C.R.,
RA Garsky V.M., Johnson R.G.;
RT "Solution structure of the cytoplasmic domain of phospholamban:
RT phosphorylation leads to a local perturbation in secondary
RT structure.";
RL Biochemistry 34:7603-7613(1995).
RN [6]
RP 3D-STRUCTURE MODELING.
RX MEDLINE=95269058; PubMed=7749920;
RA Adams P.D., Atkin I.T., Engelman D.M., Bruenger A.T.;
RT "Computational searching and mutagenesis suggest a structure for the
RT pentameric transmembrane domain of phospholamban.";
RL Nat. Struct. Biol. 2:154-162(1995).
RN [7]
RP 3D-STRUCTURE MODELING.
RX MEDLINE=98170970; PubMed=9512019;
RA Herzyk P., Hubbard R.E.;
RT "Using experimental information to produce a model of the
RT transmembrane domain of the ion channel phospholamban.";
RL Biophys. J. 74:1203-1214(1998).
RN [8]
RP FUNCTION: PHOSPHOLAMBAN HAS BEEN POSTULATED TO REGULATE THE
CC ACTIVITY OF THE CALCIUM PUMP OF CARDIAC SARCOPLASMIC RETICULUM.
CC -1- SUBUNIT: Homopentamer.
CC -1- SUBCELLULAR LOCATION: Membrane.
CC -1- TISSUE SPECIFICITY: HEART.
CC -1- PTM: PHOSPHORYLATED IN RESPONSE TO BETA-ADRENERGIC STIMULATION.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----

```

```

DR EMBL; M63603; AAA60083.1; -.
DR EMBL; M60411; AAA60109.1; -.
DR EMBL; A0177764; AAD55950.1; -.
DR EMBL; BC005269; AAH05269.1; -.
DR PIR; A40424; A40424.
DR PDB; 1PLP; 31-JUL-95.
DR PDB; 1K9N; 14-NOV-01.
DR PDB; 1KCH; 28-NOV-01.
DR PDB; 1PLN; 29-DEC-95.
DR PDB; 1PSL; 03-JUN-95.
DR Genew; HGNC:9080; PLN.
DR MIM; 172405; -.
DR InterPro; IPR005984; P_lamban.
DR Pfam; PF04272; Phospholamban; 1.
DR TIGRfams; TIGR01294; P_lamban; 1.
KW Transmembrane; Phosphorylation; Acetylation; 3D-structure.
FT DOMAIN 1 31 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 32 52 POTENTIAL.
FT MOD_RES 1 1 ACETYLATION (BY SIMILARITY).
FT MOD_RES 16 16 PHOSPHORYLATION (BY PKA) (BY SIMILARITY).
FT MOD_RES 17 17 PHOSPHORYLATION (BY CAMK) (BY SIMILARITY).
FT TURN 2 3
FT HELIX 4 14
FT TURN 15 15
SQ SEQUENCE 52 AA; 6108 MW; 0766304A76A854D3 CRC64;

Query Match 4.6%; Score 6; DB 1; Length 52;
Best Local Similarity 100.0%; Pred. No. 40;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 108 AIRRAS 113
DB 11 AIRRAS 16

RESULT 29
PPLA_MOUSE STANDARD; PRT; 52 AA.
AC P2006;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Cardiac phospholamban (PLB).
GN PLN.
OS Mus musculus (Mouse),
OS Rattus norvegicus (Rat), and
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090, 10116, 9986;
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES=Rabbit;
RX MEDLINE=91268032; PubMed=1828805;
RA Fujii J., Zazain-Herzberg A., Willard H.F., Tada M., MacLennan D.H.;
RT "Structure of the rabbit phospholamban gene, cloning of the human
RT cDNA, and assignment of the gene to human chromosome 6.";
RL J. Biol. Chem. 266:11669-11675(1991).
RN [2]
RP SEQUENCE FROM N.A.
RC SPECIES=Rabbit;
RX MEDLINE=88112222; PubMed=2962883;
RA Fujii J., Lytton J., Tada M., MacLennan D.H.;
RT "Rabbit cardiac and slow-twitch muscle express the same phospholamban
RT gene.";
RL FEBS Lett. 227:51-55(1988).
RN [3]
RP SEQUENCE FROM N.A.
RC SPECIES=Rat;
RX MEDLINE=93075183; PubMed=1445334;
RA Johns D.C., Feldman A.M.;
RT "Identification of a highly conserved region at the 5' flank of the

```

```

RT phospholamban gene.";
RL Biochem. Biophys. Res. Commun. 188:927-933(1992).
RN [4]
RP SEQUENCE FROM N.A.
RC SPECIES=Rat;
RX MEDLINE=93284726; PubMed=8508530;
RA Shanahan C.M., Weisberg P.L., Metcalfe J.C.;
RT "Isolation of gene markers of differentiated and proliferating
RT vascular smooth muscle cells.";
RL Circ. Res. 73:193-204(1993).
RN [5]
RP SEQUENCE FROM N.A.
RC SPECIES=Rat; TISSUE=Aortic smooth muscle;
RX MEDLINE=92206263; PubMed=1725098;
RA Hwang K.S., Nadal-Ginard B.;
RT "Cloning phospholamban cDNA from rat aortic smooth muscle.";
RL Adv. Exp. Med. Biol. 304:387-395(1991).
RN [6]
RP SEQUENCE FROM N.A.
RC SPECIES=Mouse;
RX MEDLINE=9308802; PubMed=1394867;
RA Kadambi V., Neumann J.C., Doetschman T., Kranias E.G.;
RT "Mouse phospholamban gene expression during development in vivo and
RT in vitro.";
RL Circ. Res. 71:1021-1030(1992).
CC -1- FUNCTION: PHOSPHOLAMBAN HAS BEEN POSTULATED TO REGULATE THE
CC ACTIVITY OF THE CALCIUM PUMP OF CARDIAC SARCOPLASMIC RETICULUM.
CC -1- SUBUNIT: Homopentamer.
CC -1- TISSUE SPECIFICITY: HEART.
CC -1- TISSUE PHOSPHORYLATED IN RESPONSE TO BETA-ADRENERGIC STIMULATION.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sb-sib.ch).
CC
CC -----
CC EMBL; S46792; AAB23706.1; -.
CC EMBL; M63600; -; NOT ANNOTATED_CDS.
CC EMBL; M63601; AAA31445.1; -.
CC EMBL; Y00761; CAA68730.1; -.
CC EMBL; L03382; AAA41848.1; -.
CC EMBL; S95853; AAB21903.1; -.
CC EMBL; S95849; AAN86727.1; -.
CC EMBL; X71068; CAA50394.1; -.
CC PIR; A49057; A49057.
CC PIR; B40424; B40424.
CC PIR; S37638; S37638.
CC HSSP; P26678; 1PLP.
CC MED; MG1:97622; PLN.
CC InterPro; IPR005984; P_lamban.
CC Pfam; PF04272; Phospholamban; 1.
CC TIGRfams; TIGR01294; P_lamban; 1.
KW Transmembrane; Phosphorylation; Acetylation.
FT DOMAIN 1 31 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 32 52 POTENTIAL.
FT MOD_RES 1 1 ACETYLATION (BY SIMILARITY).
FT MOD_RES 16 16 PHOSPHORYLATION (BY PKA) (BY SIMILARITY).
FT MOD_RES 17 17 PHOSPHORYLATION (BY CAMK) (BY SIMILARITY).
SQ SEQUENCE 52 AA; 6094 MW; 0763601F76A854D3 CRC64;

Query Match 4.6%; Score 6; DB 1; Length 52;
Best Local Similarity 100.0%; Pred. No. 40;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 108 AIRRAS 113
DB 11 AIRRAS 16

```

```

RESULT 30
ID_PPIA_PIG STANDARD; PRT; 52 AA.
AC P07473;
DT 01-APR-1988 (Rel. 07, Created)
DT 01-APR-1988 (Rel. 07, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Cardiac phospholamban (PLB).
GN PLN.
OS Sus acrofa (Pig), and
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9823, 9615;
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES=C.familiaris;
RX MEDLINE=87083954; PubMed=3793929;
RA Fujii J., Ueno A., Kitano K., Tanaka S., Kadoma M., Tada M.;
RT "Complete complementary DNA-derived amino acid sequence of canine
RT cardiac phospholamban."
RL J. Clin. Invest. 79:301-304(1987).
RN [2]
RP SEQUENCE FROM N.A.
RC SPECIES=C.familiaris; TISSUE=Heart ventricle;
RX MEDLINE=87316936; PubMed=3628007;
RA Ueda A., Kitano K., Fujii J., Kadoma M., Tada M., Tanaka S.;
RT "The cDNA sequence of the major phospholamban mRNA in canine cardiac
RT ventricular muscle."
RL Nucleic Acids Res. 15:6738-6738(1987).
RN [3]
RP SEQUENCE OF 1-45.
RC SPECIES=C.familiaris;
RX MEDLINE=8623152; PubMed=3753485;
RA Fujii J., Kadoma M., Tada M., Toda H., Sakiyama F.;
RT "Characterization of structural unit of phospholamban by amino acid
RT sequencing and electrophoretic analysis."
RL Biochem. Biophys. Res. Commun. 138:1044-1050(1986).
RN [4]
RP SEQUENCE OF 10-45.
RC SPECIES=C.familiaris;
RX MEDLINE=87008549; PubMed=3759968;
RA Simmerman H.K.B., Collins J.H., Theibert J.L., Wegener A.D.,
RA Jones L.R.;
RT "Sequence analysis of phospholamban. Identification of
RT phosphorylation sites and two major structural domains."
RL J. Biol. Chem. 261:13333-13341(1986).
RN [5]
RP PHOSPHORYLATION.
RC SPECIES=C.familiaris;
RX MEDLINE=89291905; PubMed=2544595;
RA Wegener A.D., Simmerman H.K.B., Lindemann J.P., Jones L.R.;
RT Phospholamban phosphorylation in intact ventricles. Phosphorylation
RT of serine 16 and threonine 17 in response to beta-adrenergic
RT stimulation."
RL J. Biol. Chem. 264:11468-11474(1989).
RN [6]
RP SEQUENCE FROM N.A.
RC SPECIES=Pig; TISSUE=Smooth muscle;
RX MEDLINE=90056437; PubMed=2530978;
RA Verboom H., Wuytack F., Eggermont J.A., de Jaegere S.,
RA Misaen L., Raeymaekers L., Castels R.;
RT "cDNA cloning and sequencing of phospholamban from pig stomach smooth
RT muscle."
RL Biochem. J. 262:353-356(1989).
RN [7]
RP FUNCTION: PHOSPHOLAMBAN HAS BEEN POSTULATED TO REGULATE THE
RN ACTIVITY OF THE CALCIUM PUMP OF CARDIAC SARCOPLASMIC RETICULUM.
RN [8]
RP SUBUNIT: Homopentamer.
RN [9]
RP SUBCELLULAR LOCATION: Membrane.
RN [10]
RP TISSUE SPECIFICITY: HEART.
RN [11]
RP PTM: PHOSPHORYLATED IN RESPONSE TO BETA-ADRENERGIC STIMULATION.

```

```

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M16012; AAA30884.1; -
DR EMBL; X00399; CAA58461.1; -
DR EMBL; M35393; AAC41618.1; -
DR EMBL; X15075; CAA33171.1; -
DR PIR; A29002; A29002.
DR PIR; S05540; S05540.
DR PDB; 1RJK; 06-SEP-00.
DR InterPro; IPR005984; P_lamban.
DR Pfam; PF04272; Phospholamban; 1.
DR TIGRFAMs; TIGR01294; P_lamban; 1.
KW Transmembrane; Phosphorylation; Acetylation; 3D-structure.
FT DOMAIN 1 31 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 32 52 POTENTIAL.
FT MOD RES 1 1 ACETYLATION.
FT MOD RES 16 16 PHOSPHORYLATION (BY PKA).
FT MOD RES 17 17 PHOSPHORYLATION (BY CMK1).
SQ SEQUENCE 52 AA; 6080 MW; 076361D9ADC424D3 CRC64;

Query Match 4.6%; Score 6; DB 1; Length 52;
Best Local Similarity 100.0%; Pred. No. 40;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 108 AIRRAS 113
Db 11 AIRRAS 16

RESULT 31
ID_RK33_GUITH STANDARD; PRT; 56 AA.
AC 078487;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Chloroplast 50S ribosomal protein L33.
GN RPL33.
OS Guillardia theta (Cryptomonas phi).
OC Chloroplast.
OC Eukaryota; Cryptophyta; Cryptomonadaceae; Guillardia.
OX NCBI_TaxID=55529;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99128221; PubMed=9929392;
RA Douglas S.B., Penny S.L.;
RT "The plastid genome of the cryptophyte alga, Guillardia theta:
RT complete sequence and conserved syntenic groups confirm its common
RT ancestry with red algae."
RL J. Mol. Evol. 48:236-244(1999).
RN [2]
RP SIMILARITY: BELONGS TO THE L33P FAMILY OF RIBOSOMAL PROTEINS.
RN [3]
RP THIS SWISS-PROT entry is copyright. It is produced through a collaboration
RN between the Swiss Institute of Bioinformatics and the EMBL outstation -
RN the European Bioinformatics Institute. There are no restrictions on its
RN use by non-profit institutions as long as its content is in no way
RN modified and this statement is not removed. Usage by and for commercial
RN entities requires a license agreement (See http://www.isb-sib.ch/announce/
RN or send an email to license@isb-sib.ch).
RN [4]
DR EMBL; AF041468; AAC35678.1; -
DR HAMAP; MF 00294; -; 1.
DR InterPro; IPR001705; Ribosomal_L33.
DR Pfam; PF00471; Ribosomal_L33; 1.
DR ProDom; PD002595; Ribosomal_L33; 1.
DR TIGRFAMs; TIGR01023; rplmg_dact; 1.

```

DR PROSITE; PS00582; RIBOSOMAL\_L33; 1.  
 KW Ribosomal protein; Chloroplast.  
 SQ SEQUENCE 56 AA; 6663 MW; ABDAA44411E778C1 CRC64;  
 Query Match 4.6%; Score 6; DB 1; Length 56;  
 Best Local Similarity 100.0%; Pred. No. 42;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 38 GVRIVV 43  
 DB 6 GVRIVV 11

RESULT 32  
 CD24\_MOUSE STANDARD; PRT; 76 AA.  
 ID CD24\_MOUSE STANDARD; PRT; 76 AA.  
 AC P24807; P26691;  
 DT 01-MAR-1992 (Rel. 21, Created)  
 DT 01-MAR-1992 (Rel. 21, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Signal transducer CD24 precursor (M1/69-J11d heat stable antigen) (HSA) (Nectadrin) (L1-52) (X62 heat stable antigen) (R13-A6).  
 GN CD24 OR CD24A.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC MEDLINE=90361906; PubMed=2118158;  
 RA Kay R., Taket F., Humphries R.K.;  
 RT "Expression cloning of a cDNA encoding M1/69-J11d heat-stable antigenens."; Eur. J. Immunol. 21:1039-1046(1991).  
 RL [3]  
 RN REVISIONS.  
 RP Nielsen P.J.;  
 RL Submitted (JUL-1992) to the EMBL/GenBank/DBJ databases.  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CS7BL/6 X CBA; TISSUE=Spleen;  
 RX MEDLINE=94043127; PubMed=8226859;  
 RA Wenger R.H., Rochelle J.M., Seidlin M.F., Koehler G., Nielsen P.J.;  
 RT "The heat stable antigen (mouse CD24) gene is differentially regulated but has a housekeeping promoter."; Eur. J. Biol. Chem. 268:23345-23352(1993).  
 RL [5]  
 RN SEQUENCE OF 27-53.  
 RP STRAIN=CS7BL/6;  
 RX MEDLINE=92412120; PubMed=1530634;  
 RA Hitsuoto Y., Nakano A., Onishi H., Hamada F., Saeki S., Takeuchi N.;  
 RT "Purification of the murine heat-stable antigen from erythrocytes"; Biochem. Biophys. Res. Commun. 187:773-777(1992).  
 RL [1]  
 RN FUNCTION: MAY HAVE A SPECIFIC ROLE TO PLAY IN EARLY THYMOCYTE DEVELOPMENT.  
 CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.  
 CC -1- TISSUE SPECIFICITY: IN LYMPHOID, MYELOID, AND ERYTHROID CELLS.  
 CC -1- PPM: EXTENSIVELY O-GLYCOSYLATED (BY SIMILARITY).  
 CC -1- SIMILARITY: TO OTHER MAMMALIAN SPECIES CD24.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation-CC the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; M58661; AAA39481.1; -  
 DR EMBL; X56469; CAA39841.1; -  
 DR EMBL; X72910; CAA51415.1; -  
 DR EMBL; X53825; CAA37822.1; -  
 DR PIR; A43537; A43537.  
 DR MGI; 88323; CQ24a.  
 KW Glycoprotein; GPI-anchor; Membrane; Signal; Antigen.  
 FT SIGNAL 1 26  
 FT CHAIN 27 53  
 FT PROPEP 54 76  
 FT CARBOHYD 27 27  
 FT CARBOHYD 30 30  
 FT CARBOHYD 39 39  
 FT CARBOHYD 48 48  
 FT LIPID 53 53  
 SQ SEQUENCE 76 AA; 7797 MW; 6853F121B33625EB CRC64;  
 Query Match 4.6%; Score 6; DB 1; Length 76;  
 Best Local Similarity 100.0%; Pred. No. 55;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 23 QTSVAP 28  
 DB 28 QTSVAP 33

RESULT 33  
 CD24\_RAT STANDARD; PRT; 76 AA.  
 ID CD24\_RAT STANDARD; PRT; 76 AA.  
 AC 007450;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE Signal transducer CD24 precursor (Heat stable antigen) (HSA) (Nectadrin).  
 GN CD24A.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Wistar; TISSUE=Embryonic brain;  
 RX MEDLINE=94122434; PubMed=8292828;  
 RA Shirasawa T., Akashi T., Sakamoto K., Takahashi H., Maruyama N., Hirokawa K.;  
 RT "Gene expression of CD24 core peptide molecule in developing brain and developing non-neural tissues."; Dev. Dyn. 198:1-13(1993).  
 RL [2]  
 RN SEQUENCE FROM N.A.  
 RP STRAIN=Fischer;  
 RX MEDLINE=97157759; PubMed=9004038;  
 RA Megraldo T.A., Barrandon Y.;  
 RT "CD24 (heat stable antigen, nectadrin), a novel keratinocyte differentiation marker, is preferentially expressed in areas of the hair follicle containing the colony-forming cells"; J. Cell Sci. 109:3035-3045(1996).  
 RL [1]  
 RN FUNCTION: MAY HAVE A PIVOTAL ROLE IN CELL DIFFERENTIATION. THE TRIGGERING MECHANISM OF SIGNAL TRANSDUCTION MAY BE DUE TO THE INTERACTIONS OF DIFFERENTIATING CELLS WITH THE MATRIX SUBSTRATE VIA THE CARBOHYDRATE STRUCTURE OF THE MOLECULE. IN THIS WAY, THE SIGNAL TRANSDUCER CAN PLAY VERY DIFFERENT ROLES IN DIFFERENT CELL TYPES AS A DIRECT CONSEQUENCE OF ITS GLYCOSYLATION.  
 CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.  
 CC -1- TISSUE SPECIFICITY: EXPRESSED IN THE CENTRAL NERVOUS SYSTEM, IN POSTMITOTIC CELLS OF SPINAL CORD, HINDERAIN, MIDBRAIN AND FOREBRAIN. EXPRESSED IN EPITHELIUM DURING THE DEVELOPMENT OF NON-NEURAL TISSUES. EXPRESSED IN TOOTH DEVELOPMENT, SPECIFICALLY IN

```

CC MESSENGER CELLS DIFFERENTIATING INTO ODONTOBLAST IN DENTAL
CC PAPILLA, AS WELL AS IN THE DEVELOPING EYE AND HAIR FOLLICLE.
CC -1- DEVELOPMENTAL STAGE: DETECTED IN PRIMITIVE ECTODERM, MESODERM AND
CC VENTRAL ENDODERM; DOWN-REGULATED WHEN ORGANOGENESIS IS COMPLETED.
CC -1- PM: EXTENSIVELY O-GLYCOSYLATED (BY SIMILARITY). THE CARBOHYDRATE
CC STRUCTURE MAY BE REGULATED IN A TISSUE SPECIFIC AND DEVELOPMENTAL
CC STAGE SPECIFIC MANNER.
CC -1- SIMILARITY: TO OTHER MAMMALIAN SPECIES CD24.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z11663; CAJ77731.1; -
DR EMBL; U49062; AAA91470.1; -
DR PIR; I53107; I53107.
KW Glycoprotein; GPI-anchor; Membrane; Signal; Differentiation.
FT SIGNAL 1 26
FT CHAIN 27 56
FT PROPEP 57 76
FT CARBOHYD 27 27
FT CARBOHYD 37 37
FT CARBOHYD 48 48
FT LIPID 56 56
FT SEQUENCE 76 AA; 7862 MW; 42846F70EC39D58 CRC64;
SQ
Query Match 4.6%; Score 6; DB 1; Length 76;
Best Local Similarity 100.0%; Pred. No. 55;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 23 QTSVAP 28
Db 28 QTSVAP 33
RESULT 34
PYS1_SYNY3 STANDARD; PRT; 83 AA.
AC P73202;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Phycobilisome 8.9 kDa linker polypeptide, phycocyanin-associated, rod
DE (L1-8.9/R1) (Rod capping linker protein).
GN PCPD OR SSI3093.
OS Synechocystis sp. (strain PCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
OX NCBI_TaxId=1146;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97061201; PubMed=8905231;
RA Kaneo T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
RA Miyajima N., Hirosewa M., Sugiyama M., Sasamoto S., Kimura T.,
RA Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Nanno K.,
RA Okumura S., Shimpo S., Takeuchi C., Wada T., Watanabe A.,
RA Yamada M., Yasuda M., Tabata S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium
RT Synechocystis sp. strain PCC6803. II. Sequence determination of the
RT entire genome and assignment of potential protein-coding regions.";
RL DNA Res. 3:109-136(1996).
RN [2]
RP SEQUENCE OF 1-20.
RX MEDLINE=97443974; PubMed=9298645;
RA Sazuka T., Ohara O.;
RT "Towards a proteome project of cyanobacterium Synechocystis sp.
RT strain PCC6803: linking 130 protein spots with their respective
RT genes.";
RL Electrophoresis 18:1252-1258(1997).
CC -1- SIMILARITY: TO OTHER PHYCOBILISOME LINKER PROTEINS.

```

```

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D90904; BAA17228.1; -
DR PIR; S75314; S75314.
DR InterPro; IPR001685; CpcD-like.
DR Pfam; PF01383; CpcD; 1.
DR ProDom; PD002828; CpcD-like C; 1.
KW Phycobilisome; Photosynthesis; Complete proteome.
SQ SEQUENCE 83 AA; 9322 MW; 95F1570CFCAD561 CRC64;
SQ
Query Match 4.6%; Score 6; DB 1; Length 83;
Best Local Similarity 100.0%; Pred. No. 60;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 73 SRLGCT 78
Db 63 SRLGCT 68
RESULT 35
YAES_YEAST STANDARD; PRT; 102 AA.
ID YAES_YEAST
AC P39725;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Hypothetical 11.6 kDa protein in ACS1-GC33 intergenic region.
GN YAL045C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycaceae.
OX NCBI_TaxId=4932;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=S288C / AB972;
RX MEDLINE=95249563; PubMed=7711988;
RA Bussey H., Kaback D.B., Zhong W., Vo D.T., Clark M.W., Fortin N.,
RA Hall J., Quellerie B.F.F., Keng T., Barton A.B., Su Y., Davies C.K.,
RA Storms R.K.;
RT "The nucleotide sequence of chromosome I from Saccharomyces
RT cerevisiae.";
RT Proc. Natl. Acad. Sci. U.S.A. 92:3809-3813(1995).
RL -1- SIMILARITY: TO E. COLI BOLA.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U12980; AAC04986.1; -
DR PIR; S51974; S51974.
DR SGD; S0000043; YAL045C.
KW Hypothetical protein.
SQ SEQUENCE 102 AA; 11648 MW; 06B50906B90D5509 CRC64;
SQ
Query Match 4.6%; Score 6; DB 1; Length 102;
Best Local Similarity 100.0%; Pred. No. 71;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 39 VRIIVE 44
Db 55 VRIIVE 60

```

```

RESULT 36
RS6_CHLMTU STANDARD; PRT; 112 AA.
ID RS6_CHLMTU
AC O9PLC1;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE 30S ribosomal protein S6.
RPSF OR TC0184.
OS Chlamydia muridarum.
OC Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydia.
NCBI_TaxID=83560;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=MOPH / N199;
RX MEDLINE=20150255; PubMed=10684935;
RA Read T.D., Brunham R.C., Shen C., Gail S.R., Heidelberg J.F.,
RA White O., Hickey E.K., Peterson J., Utecherback T., Berry K., Baas S.,
RA Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,
RA Gwinn M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,
RA Eisen J., Fraser C.M.,
RT "Genome sequences of Chlamydia trachomatis MOPH and Chlamydia
RT pneumoniae AR39."
RT Nucleic Acids Res. 28:1397-1406(2000).
CC -1- FUNCTION: BINDS TOGETHER WITH S18 TO 16S RIBOSOMAL RNA
CC (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE S6P FAMILY OF RIBOSOMAL PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE002285; AAF39058.1; -.
DR PIR; D81732; D81732.
DR TIGR; TC0184; -.
DR HAMAP; MF_00360; -.
DR InterPro; IPR000529; Ribosomal_S6.
DR Pfam; PF01250; Ribosomal_S6; 1.
DR ProDom; PD003809; Ribosomal_S6; 1.
DR TIGRFAMs; TIGR00166; S6; 1.
DR PROSITE; PS01048; RIBOSOMAL_S6; FALSE NEG.
KW Ribosomal protein; rRNA-binding; Complete proteome.
SQ SEQUENCE 112 AA; 12923 MW; 1A37A64AA72355C5 CRC64;

Query Match 4.6%; Score 6; DB 1; Length 112;
Best Local Similarity 100.0%; Pred. No. 77;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 59 ASAYKE 64
Db 98 ASAYKE 103

RESULT 37
RS6_CHLTR STANDARD; PRT; 112 AA.
ID RS6_CHLTR
AC O84807;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE 30S ribosomal protein S6.
RPSF OR RS6 OR CR801.
OS Chlamydia trachomatis.
OC Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydia.
NCBI_TaxID=813;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=D/UW-3/Cx;

```

```

RX MEDLINE=99000809; PubMed=9784136;
RA Stephens R.S., Kaiman S., Lammel C.J., Fan J., Marathe R., Aravind L.,
RA Mitchell W.P., Olinger L., Tatusov R.L., Zhao Q., Koonin E.V.,
RA Davis R.W.,
RT "Genome sequence of an obligate intracellular pathogen of humans:
RT Chlamydia trachomatis."
RL Science 282:754-759(1998).
CC -1- FUNCTION: BINDS TOGETHER WITH S18 TO 16S RIBOSOMAL RNA
CC (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE S6P FAMILY OF RIBOSOMAL PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE001352; AAC68396.1; -.
DR PIR; G71469; G71469.
DR PHCI; 2DPAGE; O84807; -.
DR HAMAP; MF_00360; -.
DR InterPro; IPR000529; Ribosomal_S6.
DR Pfam; PF01250; Ribosomal_S6; 1.
DR ProDom; PD003809; Ribosomal_S6; 1.
DR TIGRFAMs; TIGR00166; S6; 1.
DR PROSITE; PS01048; RIBOSOMAL_S6; FALSE NEG.
KW Ribosomal protein; rRNA-binding; Complete proteome.
SQ SEQUENCE 112 AA; 12907 MW; 1A37BFD394355C5 CRC64;

Query Match 4.6%; Score 6; DB 1; Length 112;
Best Local Similarity 100.0%; Pred. No. 77;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 59 ASAYKE 64
Db 98 ASAYKE 103

RESULT 38
TAT_SIVGB STANDARD; PRT; 115 AA.
ID TAT_SIVGB
AC P22384;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE TAT protein (Transactivating regulatory protein).
GN TAT.
OS Simian immunodeficiency virus (isolate GB1).
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.
NCBI_TaxID=11732;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=90015169; PubMed=2797181;
RA Tsujimoto H., Hasegawa A., Maki N., Fukasawa M., Miura T., Speidel S.,
RA Cooper R.W., Moriyama E.N., Gojobori T., Hayami M.,
RT "Sequence of a novel simian immunodeficiency virus from a wild-caught
RT African mandrill."
RL Nature 341:539-541(1998).
RN [2]
RP MUTAGENESIS OF CYS-37, AND INTERACTION WITH CYCLIN T.
RX MEDLINE=99292873; PubMed=10364329;
RA Bientz P.D., Grdina T.A., Bogerd H.P., Gullen B.R.,
RT "Analysis of the effect of natural sequence variation in Tat and in
RT cyclin T on the formation and RNA binding properties of Tat-cyclin T
RT complexes."
RL J. Virol. 73:5777-5786(1999).
CC -1- FUNCTION: TRANSCRIPTIONAL REGULATOR THAT ACTS BY BINDING TO THE
CC TRANS-ACTIVATING SEQUENCE (TAR) RNA ELEMENT AND
CC ACTIVATES TRANSCRIPTION INITIATION AND/OR ELONGATION FROM THE LTR
CC PROMOTER.
CC -1- SUBUNIT: BINDS CYCLIN T1.

```

```
CC -1- SUBCELLULAR LOCATION: Nuclear; nucleolar.
CC -1- MISCELLANEOUS: THIS IS AN AFRICAN MANDRILL ISOLATE.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
```

EMBL: M27470; AAB49572.1; -  
DR InterPro: IPR001831; HIV\_Tat.  
DR Pfam; PF00539; Tat; 1.  
DR PRINTS; PR00055; HIVTATDOMAIN.  
KW Transcription regulation; Activator; RNA-binding; Nuclear protein;  
KM AIDS.  
FT MDTAGN 37 37 C-S: LOSS OF BINDING TO CYCLIN T.  
SQ SEQUENCE 115 AA; 1348 MW; AF220B634A585FDC CRC64;

Query Match 4.6%; Score 6; DB 1; Length 115;  
Best Local Similarity 100.0%; Pred. No. 79;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 107 FAIRRA 112  
Db 83 FAIRRA 88

RESULT 39  
TAT\_SIVAM STANDARD; PRT; 116 AA.  
ID TAT\_SIVAM  
AC P36340;  
DT 01-JUN-1994 (Rel. 29, Created)  
DT 01-FEB-1996 (Rel. 33, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE TAT protein (Transactivating regulatory protein).  
GN TAT.  
OS Simian immunodeficiency virus (isolate African mandrill) (SIV).  
OX Viruses; Retroviruses; Retroviridae; Lentiviruses.  
NCBI\_TaxID=36378;  
RN (1)  
RP SEQUENCE FROM N.A.  
RX MEDLINE=92352315; PubMed=1642547;  
RA Sakai H., Sakuragi J., Sakuragi S., Shibata R., Hayami M.,  
RA Ishimoto A., Adachi A.;  
RT "Genetic characterization of simian immunodeficiency virus isolated  
RT from an African mandrill.";  
RL Arch. Virol. 125:1-14(1992).  
CC -1- FUNCTION: TRANSCRIPTIONAL REGULATOR THAT ACTS BY BINDING TO THE  
CC TRANS-ACTIVATING RESPONSIVE SEQUENCE (TAR) RNA ELEMENT AND  
CC ACTIVATES TRANSCRIPTION INITIATION AND/OR ELONGATION FROM THE LTR  
CC PROMOTER.  
CC -1- SUBUNIT: BINDS CYCLIN T1 (BY SIMILARITY).  
CC -1- SUBCELLULAR LOCATION: Nuclear; nucleolar.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL: M62884; AAA47725.1; -  
DR EMBL; M62884; AAA47725.1; JOINED.  
DR FIR; A48344; A48344.  
DR InterPro; IPR001831; HIV\_Tat.  
DR Pfam; PF00539; Tat; 1.  
DR PRINTS; PR00055; HIVTATDOMAIN.  
KW Transcription regulation; Activator; RNA-binding; Nuclear protein;  
KM AIDS.  
SQ SEQUENCE 116 AA; 13373 MW; 8FDB3D2485CFEFD6 CRC64;

Query Match 4.6%; Score 6; DB 1; Length 116;  
Best Local Similarity 100.0%; Pred. No. 80;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 107 FAIRRA 112  
Db 83 FAIRRA 88

RESULT 40  
R18E\_PYRFU STANDARD; PRT; 120 AA.  
ID R18E\_PYRFU  
AC Q8U055;  
DT 28-FEB-2003 (Rel. 41, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE 50S ribosomal protein L18e.  
GN RPL18E OR P1646.  
OS Pyrococcus furiosus.  
OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;  
OC Pyrococcus.  
OX NCBI\_TaxID=2261;  
RN (1)  
RP SEQUENCE FROM N.A.  
RC STRAIN=Vc1 / DSM 3638 / ATCC 43587 / JCM 8422;  
RA Weiss R.B., Dunn D.M., Robb P.T., Brown J.R.;  
RT "The complete sequence of the Pyrococcus furiosus genome.";  
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.  
CC -1- SIMILARITY: BELONGS TO THE L18E FAMILY OF RIBOSOMAL PROTEINS.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL: AB010263; AAL81770.1; -  
DR HAMAP; MF\_00329; -; 1.  
DR InterPro; IPR001196; Ribosomal\_L15.  
DR InterPro; IPR000039; Ribosomal\_L18e.  
DR Pfam; PF00256; L15; 1.  
DR PROSITE; PS01106; RIBOSOMAL\_L18E; FALSE\_NEG.  
KW Ribosomal protein; Complete proteome.  
SQ SEQUENCE 120 AA; 13732 MW; 60DD80150762E89 CRC64;

Query Match 4.6%; Score 6; DB 1; Length 120;  
Best Local Similarity 100.0%; Pred. No. 82;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 36 GSGVRI 41  
Db 113 GSGVRI 118

RESULT 41  
R118\_BACHD STANDARD; PRT; 120 AA.  
ID R118\_BACHD  
AC Q92978; Q92978;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE 50S ribosomal protein L18.  
GN RPLR OR BH0150.  
OS Bacillus halodurans.  
OC Bacillus; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
OX NCBI\_TaxID=86665;  
RN (1)  
RP SEQUENCE FROM N.A.  
RC STRAIN=C-125 / JCM 9153;  
RX MEDLINE=9209008; PubMed=10192928;



RA Takami H., Takaki Y., Nakasone K., Hirama C., Inoue A., Horikoshi K.;  
 RT "Sequence analysis of a 32-kb region including the major ribosomal  
 RT protein gene clusters from alkaliphilic *Bacillus* sp. strain C-125.";  
 RL Biocet. Biotechnol. Biochem. 63:452-455(1999).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C-125 / JCM 9153;  
 RX MEDLINE=20512582; PubMed=11058132;  
 RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,  
 RA Fuji F., Hirama C., Nakamura Y., Ogasawara N., Kohara S.,  
 RA Horikoshi K.;  
 RT "Complete genome sequence of the alkaliphilic bacterium *Bacillus*  
 RT halodurans and genomic sequence comparison with *Bacillus subtilis*.";  
 RL Nucleic Acids Res. 28:4317-4331(2000).  
 CC -1- FUNCTION: THIS IS ONE OF 3 PROTEINS THAT MEDIATE THE ATTACHMENT OF  
 CC THE 5S RNA INTO THE LARGE RIBOSOMAL SUBUNIT (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE L18P FAMILY OF RIBOSOMAL PROTEINS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: AB017508; BAA75287.1; -  
 DR EMBL: AP001507; BAB03669.1; -  
 DR PIR: T44399; T44399.  
 DR InterPro: IPR004389; Ribosomal\_L18bac.  
 DR InterPro: IPR005484; Ribosomal\_L18p.  
 DR Pfam: PF00861; Ribosomal\_L18p; 1.  
 DR ProDom: PD001394; Ribosomal\_L18bac; 1.  
 DR TRIGRAMS: TIGR00060; L18 bac; 1.  
 KW Ribosomal protein; rRNA-binding; Complete proteome.  
 SQ SEQUENCE 120 AA; 1321 MW; 16792364B5B1B1F1 CRC64;  
 QY  
 Db 93 KLENGG 98  
 64 KLENGG 69  
 Query Match 4.6%; Score 6; DB 1; Length 120;  
 Best Local Similarity 100.0%; Pred. No. 82;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 RESULT 42  
 CYCP\_RHORU STANDARD; PRT; 126 AA.  
 ID CYCP\_RHORU  
 AC P00144;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 21-JUL-1986 (Rel. 01, Last sequence update)  
 DT 01-MAY-1992 (Rel. 22, Last annotation update)  
 DE Cytochrome c.  
 OS *Rhodospirillum rubrum*.  
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhodospirillales;  
 OC Rhodospirillaceae; Rhodospirillum.  
 OX NCBI\_TaxID=1085;  
 RN [1]  
 RP SEQUENCE.  
 RC STRAIN=ATCC 11170 / SI;  
 RX MEDLINE=76069185; PubMed=172499;  
 RA Meyer T.E., Ambler R.P., Bartsch R.G., Kamen M.D.;  
 RT "Amino acid sequence of cytochrome c' from the purple photosynthetic  
 RT bacterium *Rhodospirillum rubrum* SI.";  
 RL J. Biol. Chem. 250:8416-8421(1975).  
 RN [2]  
 RP X-RAY CRYSTALLOGRAPHY (2.8 ANGSTROMS).  
 RX MEDLINE=92268030; PubMed=1316891;  
 RA Yasui M., Harada S., Kai Y., Kasai N., Kusunoki M., Matsura Y.;  
 RT "Three-dimensional structure of ferricytochrome c' from  
 RT *Rhodospirillum rubrum* at 2.8-A resolution.";  
 RL J. Biochem. 111:317-324(1992).

CC -1- FUNCTION: CYTOCHROME C' IS THE MOST WIDELY OCCURRING BACTERIAL  
 CC C-TYPE CYTOCHROME. CYTOCHROMES C' ARE HIGH-SPIN PROTEINS AND THE  
 CC HEME HAS NO SIXTH LIGAND. THEIR EXACT FUNCTION IS NOT KNOWN.  
 CC -1- SUBUNIT: Homodimer.  
 DR PIR: A00137; CQPCR.  
 DR HSSP: P00147; ICPR.  
 DR InterPro: IPR002321; CyC\_CIT.  
 DR InterPro: IPR000345; CyC\_heme\_bind.  
 DR Pfam: PF01322; Cytochrome\_C\_2; 1.  
 DR PRINTS: PR00608; CYTOCHROME\_CIT.  
 DR ProDom: PD003828; CyC\_CIT; 1.  
 DR PROSITE: PS00150; CYTOCHROME\_C; 1.  
 KW Electron transport; Heme.  
 FT BINDING 116 116 HEME (COVALENT).  
 FT BINDING 119 119 HEME (COVALENT).  
 FT METAL 120 120 IRON (HEME AXIAL LIGAND).  
 SQ SEQUENCE 126 AA; 13117 MW; 3841D67A62B74BE CRC64;  
 QY  
 Db 56 LEHSA 61  
 91 LEHSA 96  
 Query Match 4.6%; Score 6; DB 1; Length 126;  
 Best Local Similarity 100.0%; Pred. No. 86;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 RESULT 43  
 PAND\_L18IN STANDARD; PRT; 127 AA.  
 ID PAND\_L18IN  
 AC Q92AAB;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Aspartate 1-decarboxylase precursor (EC 4.1.1.11) (Aspartate alpha-  
 DE decarboxylase).  
 GN PAND OR LIN2014.  
 OS *Listeria innocua*.  
 OC Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.  
 OX NCBI\_TaxID=1642;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Clp 1126 / Serovar 6a;  
 RX MEDLINE=21537279; PubMed=11679669;  
 RA Glaser P., Frangeul L., Buchrieser C., Rusnok C., Amend A.,  
 RA Baquero F., Berche P., Bloeker H., Brandt P., Chakraborty T.,  
 RA Charbit A., Chetani F., Couve E., de Daruvar A., Deboux P.,  
 RA Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussauget O.,  
 RA Ertlan K.-D., Fsihi H., Garcia-del Portillo F., Garrido P.,  
 RA Gautier L., Goebel W., Gomez-Lopez N., Hain T., Haut J., Jackson D.,  
 RA Jones L.-M., Kaerst U., Kreft J., Kuhn M., Kunst F., Kurapat G.,  
 RA Meduno E., Maitournam A., Mata Vicente J., Ng E., Nedjari H.,  
 RA Nordstedt G., Novella S., de Pablos B., Perez-Diaz J.-C., Purcell R.,  
 RA Remmel B., Rose M., Schlueter T., Simoes N., Tietze A.,  
 RA Vazquez-Boland J.-A., Voos H., Wehland J., Cossart P.;  
 RT "Comparative genomics of *Listeria* species.";  
 RL Science 294:849-852(2001).  
 CC -1- CATALYTIC ACTIVITY: L-aspartate = beta-alanine + CO(2).  
 CC -1- COFACTOR: Pyruvoyl group (By similarity).  
 CC -1- PATHWAY: Pantothenate biosynthesis; second branch.  
 CC -1- SIMILARITY: BELONGS TO THE PAND FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: AL596170; CAC97244.1; -  
 DR PIR: AD1684; AD1684.  
 DR ListIDst: LIN02014; -



```

DR HAMAP; MF_00446; -, 1.
DR InterPro; IPR003190; Asp_decarbox.
DR Pfam; PF02261; Asp_decarbox; 1.
DR ProDom; PD009294; Asp_decarbox; 1.
DR TIGRPFAM; TIGR00223; Pand; 1.
DR Pantothenate biosynthesis; Lyase; Decarboxylase; Pyruvate; Zymogen;
KM Complete proteome.
FT CHAIN 1 24 ASPARTATE 1-DECARBOXYLASE BETA CHAIN
FT (BY SIMILARITY).
FT CHAIN 25 127 ASPARTATE 1-DECARBOXYLASE ALPHA CHAIN
FT (BY SIMILARITY).
FT MOD_RES 25 25 (BY SIMILARITY).
FT CONVERTED TO A PYRUVYL GROUP
FT (BY SIMILARITY).
SQ SEQUENCE 127 AA; 13931 MW; C8817923C0A58675 CRC64;

Query Match
Best Local Similarity 100.0%; Pred. No. 86;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 34 EPGSGV 39
Db 63 EPGSGV 68

RESULT 44
PAND_LISMO STANDARD; PRT; 127 AA.
AC 08Y603;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Aspartate 1-decarboxylase precursor (EC 4.1.1.11) (Aspartate alpha-
DE decarboxylase).
DE PAND OR ILM01900.
OS Listeria monocytogenes.
OC Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.
OX NCBI_TaxID=1639;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=EGD-e / Serovar 1/2a;
RX MEDLINE=21537279; PubMed=11679669;
RA Glauber P., Frangeul L., Buchrieser C., Rusniok C., Amend A.,
RA Baquero F., Barthelemy P., Bloeker H., Brandt P., Chakraborty T.,
RA Chabot A., Chetoui F., Couve E., de Darvar A., Denoux P.,
RA Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussauget O.,
RA Entian K.-D., Fathi H., Garcia-del Portillo F., Garrido P.,
RA Gautier L., Geibel W., Gomez-Lopez N., Hain T., Haut J., Jackson D.,
RA Jones L.-M., Kaerst U., Krell T., Kunz M., Kunst F., Kurapkhat G.,
RA Madueno E., Maitournam A., Mata Vicente J., Ng E., Nedjari H.,
RA Nordstok G., Novella S., de Pablo B., Perez-Diaz J.-C., Purcell R.,
RA Remmel B., Rose M., Schueter T., Simoes N., Tierrez A.,
RA Valquez-Boland J.-A., Voss H., Wehland J., Cossart P.;
RA "Comparative genomes of Listeria species.";
RL Science 294:849-852(2001).
CC -1- CATALYTIC ACTIVITY: L-aspartate = beta-alanine + CO(2).
CC -1- COFACTOR: Pyruvyl group (By similarity).
CC -1- PATHWAY: Pantothenate biosynthesis; second branch.
CC -1- SIMILARITY: BELONGS TO THE PAND FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AL591981; CAC39978.1; -
CC PIR; AD1312; AD1312.
CC Listlist; ILM01900; -
CC HAMAP; MF_00446; -, 1.
DR InterPro; IPR003190; Asp_decarbox.
DR Pfam; PF02261; Asp_decarbox; 1.

```

```

DR ProDom; PD009294; Asp_decarbox; 1.
DR TIGRPFAM; TIGR00223; Pand; 1.
DR Pantothenate biosynthesis; Lyase; Decarboxylase; Pyruvate; Zymogen;
KM Complete proteome.
FT CHAIN 1 24 ASPARTATE 1-DECARBOXYLASE BETA CHAIN
FT (BY SIMILARITY).
FT CHAIN 25 127 ASPARTATE 1-DECARBOXYLASE ALPHA CHAIN
FT (BY SIMILARITY).
FT MOD_RES 25 25 (BY SIMILARITY).
FT CONVERTED TO A PYRUVYL GROUP
FT (BY SIMILARITY).
SQ SEQUENCE 127 AA; 13959 MW; D6619733C0A58675 CRC64;

Query Match
Best Local Similarity 100.0%; Pred. No. 86;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 34 EPGSGV 39
Db 63 EPGSGV 68

RESULT 45
C066_HYACE STANDARD; PRT; 129 AA.
AC P45590;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Larval/pupal rigid cuticle protein 66 precursor (HCCP66).
DE C066.
OS Hyalophora cecropia (Cecropia moth).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Lepidoptera; Glossata; Ditypsia; Bombycoidea;
OC Saturniidae; Saturniinae; Attacini; Hyalophora.
OX NCBI_TaxID=7123;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 18-43 AND 78-108.
RC TISSUE=Epidermis;
RX MEDLINE=94297581; PubMed=8025561;
RA Lampe D.J., Willis J.H.;
RT "Characterization of a cDNA and gene encoding a cuticular protein
RT from rigid cuticles of the giant silkworm, Hyalophora cecropia.";
RL Insect Biochem. Mol. Biol. 24:419-435(1994).
CC -1- FUNCTION: COMPONENT OF THE RIGID CUTICLE OF THE LARVA AND PUPA OF
CC TENEBRIO MOLLITOR.
CC -1- TISSUE SPECIFICITY: EXPRESSED IN LARVAL WING DISC, FOREWING DISC,
CC DIAPAUSING WING, ADULT WING, AND IN VERY LOW AMOUNTS IN FAT BODY
CC AND TESTES.
CC -1- DEVELOPMENTAL STAGE: MOSTLY PRESENT IN YOUNG PUPAL WINGS AND
CC LARVAL TUBERCLES. FAINTLY PRESENT IN YOUNGER THAN 24H OLD PUPAL
CC FOREWING AND HINDWING. WEAKLY DETECTED IN HINDWING AND TUBERCLES.
CC -1- SIMILARITY: Contains 1 cuticle consensus domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; L13971; AAC37204.1; -
CC InterPro; IPR000618; Insect_cuticle.
DR Pfam; PF00379; Chitin_bind_4; 1.
DR PROSITE; PS00233; CUTICLE; 1.
DR structural protein; Cuticle; signal.
FT CHAIN 1 17 SIGNAL
FT SIGNAL 1 17
FT CHAIN 18 129 LARVAL/PUPAL RIGID CUTICLE PROTEIN 66.
FT CONFLICT 105 105 Y -> F (IN REF. 1; AA SEQUENCE).
FT CONFLICT 108 108 S -> F (IN REF. 1; AA SEQUENCE).
SQ SEQUENCE 129 AA; 13877 MW; 8914FEC3F94DDC6 CRC64;

Query Match
4.6%; Score 6; DB 1; Length 129;

```



Db 109 AIRRAS 114

# RESULT 48

RL28\_MOUSE STANDARD; PRT; 136 AA.

AC P41105;  
DT 01-FEB-1995 (Rel. 31, Created)  
DT 01-FEB-1995 (Rel. 31, Last sequence update)  
DT 01-NOV-1995 (Rel. 32, Last annotation update)  
DE 60S ribosomal protein L28.  
GN RPL28.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;

RC [1]  
RP SEQUENCE FROM N.A.

RC TISSUE=Testis;  
RX MEDLINE=94252590; PubMed=7515017;

RA Burke P.S., Lim E., Lin C.S., Wolgemuth D.J.;  
RT "Sequence and expression of a cDNA encoding the mouse homologue of  
the rat ribosomal protein L28.";

RL Gene 142:315-316(1994).

CC -1- SIMILARITY: BELONGS TO THE L28E FAMILY OF RIBOSOMAL PROTEINS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; X74856; CA552848.1; -.  
DR PIR; I48738; I48738.

DR MGD; MGI:101839; RPI28.  
DR InterPro; IPR002672; Ribosomal\_L28e.

DR Pfam; PF01778; Ribosomal\_L28e; 1.  
DR Prodom; PD010767; Ribosomal\_L28e; 1.

DR Ribosomal protein.  
FT INIT MET 0 BY SIMILARITY.

FT SEQUENCE 136 AA; 15602 MW; 6551A0685F9B3742 CRC64;

Query Match 4.6%; Score 6; DB 1; Length 136;  
Best Local Similarity 100.0%; Pred. No. 92;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 108 AIRRAS 113

Db 109 AIRRAS 114

RESULT 49  
RL28\_MOUSE STANDARD; PRT; 136 AA.

AC P17702;  
DT 01-AUG-1990 (Rel. 15, Created)  
DT 01-FEB-1991 (Rel. 17, Last sequence update)  
DT 01-NOV-1997 (Rel. 35, Last annotation update)  
DE 60S ribosomal protein L28.  
GN RPL28.

OS Rattus norvegicus (Rat).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OX NCBI\_TaxID=10116;

RC [1]  
RP SEQUENCE FROM N.A. AND PARTIAL SEQUENCE.

RC STRAIN=Sprague-Dawley; TISSUE=Liver;  
RX MEDLINE=91002678; PubMed=2207170;

RA Wool I.G., Chan Y.-L., Paz V., Olivera J.;  
RT "The primary structure of rat ribosomal proteins: the amino acid

RT sequences of L27a and L28 and corrections in the sequences of S4 and  
RT S12.";

RL Biochim. Biophys. Acta 1050:69-73(1990).

CC -1- SIMILARITY: BELONGS TO THE L28E FAMILY OF RIBOSOMAL PROTEINS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; X52619; CA36846.1; -.

DR PIR; S13072; RSRT28.  
DR InterPro; IPR002672; Ribosomal\_L28e.

DR Pfam; PF01778; Ribosomal\_L28e; 1.  
DR Prodom; PD010767; Ribosomal\_L28e; 1.

DR Ribosomal protein.  
FT INIT MET 0

FT SEQUENCE 136 AA; 15717 MW; 0237BD012ACB6247 CRC64;

Query Match 4.6%; Score 6; DB 1; Length 136;  
Best Local Similarity 100.0%; Pred. No. 92;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 108 AIRRAS 113

Db 109 AIRRAS 114

RESULT 50  
ADF3\_ARATH STANDARD; PRT; 139 AA.

AC Q9ZSK4;  
DT 28-FEB-2003 (Rel. 41, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Actin-depolymerizing factor 3 (ADF 3) (ADF3).

GN ADF3 OR AT5G59880 OR MN10.12.

OS Arabidopsis thaliana (Mouse-ear cress).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eucosids II; Brassicales; Brassicaceae; Arabidopsis.

OX NCBI\_TaxID=3702;

RN [1]  
RP SEQUENCE FROM N.A.

RC STRAIN=cv. Columbia;  
RX MEDLINE=21307188; PubMed=11414611;

RA Dong C.-H., Koest B., Xia G., Chua N.-H.;  
RT "Molecular identification and characterization of the Arabidopsis  
ATADF1, ATADF5 and ATADF6 genes.";

RT Plant Mol. Biol. 45:517-527(2001).

RL [2]  
RN SEQUENCE FROM N.A.

RC STRAIN=cv. Columbia;  
RX MEDLINE=99087489; PubMed=9872454;

RA Nakamura Y., Sato S., Asamizu E., Kaneko T., Kotani H., Miyajima N.,  
RA Tabata S.;

RT "Structural analysis of Arabidopsis thaliana chromosome 5. VII.  
RT Sequence features of the regions of 1,013,767 bp covered by sixteen  
RT physically assigned P1 and TAC clones.";

RL DNA Res. 5:297-308(1998).

RN [3]  
RN SEQUENCE FROM N.A.

RC STRAIN=cv. Columbia;  
RA Shinozaki K., Davis R.W., Becker J.R., Theologis A.;

RT "Riken Arabidopsis full length cDNA clones (RAFLs) sequenced by the  
RT SSP consortium (Salk/Stanford/IGEC).";

RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.

CC -1- FUNCTION: ACTIN-DEPOLYMERIZING PROTEIN. SEVERAL ACTIN FILAMENTS  
(F-ACTIN) AND BINDS TO ACTIN MONOMERS (BY SIMILARITY).  
CC -1- SIMILARITY: BELONGS TO THE ACTIN-BINDING PROTEINS ADF FAMILY.

-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
DR EMBL; AF102821; AAD09109.1; -  
DR EMBL; AB015475; BAB08356.1; -  
DR EMBL; AF360169; AAK25879.1; -  
DR HSSP; Q39250; 1F7S.  
DR InterPro; IPR002108; Actbind\_cofin.  
DR Pfam; PF00241; cofilin\_ADP; 1.  
DR ProDom; PD002129; Actbind\_cofin; 1.  
DR SMART; SW00102; ADP; 1.  
DR PROSITE; PS00325; ACTIN\_DEPOLYMERIZING; 1.  
KW Actin-binding; Multigene family.  
FT DOMAIN 92 111 ACTIN-BINDING (POTENTIAL).  
SQ SEQUENCE 139 AA; 15922 MW; D3325AA71DD0102E CRC64;

Query Match 4.6%; Score 6; DB 1; Length 139;  
Best Local Similarity 100.0%; Pred. No. 93;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 19 GEPGQT 24  
Db 47 GEPGQT 52

Search completed: December 15, 2003, 17:06:00  
Job time : 20 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 15, 2003, 17:03:21 / Search time 34 Seconds  
(without alignments)  
994.262 Million cell updates/sec

Title: US-09-925-301-966

Perfect score: 131

Sequence: 1 AAVHTRKQGPAPAPAMSGE.....ASNGTLEKITSRPPCVIL 131

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 830525 seqs, 258052604 residues

Word size : 0

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 100 summaries

Database :

SPTREMBL\_23:\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp Vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_virus:\*  
16: sp\_bacteriap:\*  
17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	length	DB ID	Description
1	115	87.8	115	4 Q9BRT3	Q9btr3 homo sapien
2	40	30.5	115	11 Q9CQ86	Q9cqb6 mus musculu
3	9	6.9	158	4 Q9Y4Y4	Q9y4y4 homo sapien
4	9	6.9	158	4 Q9Y4Y4	Q9y4y4 homo sapien
5	9	6.9	232	10 Q8S227	Q8s227 oryza sativ
6	8	6.1	163	2 Q50044	Q50044 mycobacteri
7	8	6.1	309	17 Q9HK29	Q9hk29 thermoplasm
8	8	6.1	563	1 Q8NKN0	Q8nkn0 uncultured
9	7	5.3	106	12 Q8DY75	Q8dy75 elmltan aden
10	7	5.3	106	12 Q8BEV7	Q8bev7 human adeno
11	7	5.3	121	16 Q8BVJ1	Q8bvj1 mycoplasma
12	7	5.3	159	16 Q93RX6	Q93rx6 streptomyce
13	7	5.3	161	16 Q8F207	Q8f207 leptospira
14	7	5.3	202	16 Q8BD64	Q8bd64 vibrio vuln
15	7	5.3	207	16 Q8BSN0	Q8bsn0 ocaerobacil
16	7	5.3	209	11 Q8VE68	Q8ve68 mus musculu

17	7	5.3	209	11 Q9D7B7	Q9d7b7 mus musculu
18	7	5.3	212	15 Q8AQPI	Q8aqpi human immun
19	7	5.3	233	16 Q8ZJ20	Q8zj20 salmonella
20	7	5.3	243	2 Q8VMA4	Q8vma4 streptomyce
21	7	5.3	258	4 Q9NSY0	Q9nsy0 homo sapien
22	7	5.3	280	16 Q34774	Q34774 bacillus su
23	7	5.3	287	16 Q9RJX4	Q9rjx4 streptomyce
24	7	5.3	293	4 Q8WYS5	Q8wys5 homo sapien
25	7	5.3	294	4 Q8NXC8	Q8nxc8 homo sapien
26	7	5.3	306	16 Q9CL61	Q9cl61 pasteurella
27	7	5.3	307	16 Q9JRG2	Q9jrg2 neisseria m
28	7	5.3	319	10 Q9ZWT8	Q9zwt8 nicotiana s
29	7	5.3	322	17 Q8TYL1	Q8tyl1 methanopyru
30	7	5.3	336	10 Q967W8	Q967w8 atropa bell
31	7	5.3	336	13 Q9W6K9	Q9w6k9 odontaspis
32	7	5.3	336	13 Q9W6L1	Q9w6l1 cerotrhinus
33	7	5.3	337	16 Q8XR82	Q8xr82 raietonia s
34	7	5.3	338	10 Q9XJ41	Q9xj41 hyoscyamus
35	7	5.3	340	10 Q9XJ42	Q9xj42 atropa bell
36	7	5.3	353	10 Q9ZWT9	Q9zwt9 nicotiana s
37	7	5.3	354	16 Q8DSE4	Q8dse4 streptococc
38	7	5.3	371	10 Q93X04	Q93x04 nicotiana a
39	7	5.3	372	11 Q8K393	Q8k393 mus musculu
40	7	5.3	373	17 Q9HMG7	Q9hmg7 halobacteri
41	7	5.3	378	11 Q9CZY7	Q9czy7 mus musculu
42	7	5.3	378	11 Q8CAZ9	Q8caz9 mus musculu
43	7	5.3	378	11 Q8C227	Q8c227 mus musculu
44	7	5.3	388	10 Q93X05	Q93x05 nicotiana a
45	7	5.3	389	10 Q9RJ00	Q9rj00 streptomyce
46	7	5.3	418	5 Q9VHM3	Q9vhm3 drosophila
47	7	5.3	421	17 Q97W60	Q97w60 sulfolobus
48	7	5.3	430	10 Q9ZWT7	Q9zwt7 nicotiana s
49	7	5.3	438	16 Q9KZ58	Q9kz58 streptomyce
50	7	5.3	454	16 Q8YX47	Q8yx47 arabacena sp
51	7	5.3	464	16 Q8G0S5	Q8g0s5 bruceella su
52	7	5.3	467	17 Q9Y9F2	Q9y9f2 aeropyrum p
53	7	5.3	473	5 Q94416	Q94416 caenorhabdi
54	7	5.3	483	16 Q8YH32	Q8yh32 bruceella me
55	7	5.3	488	16 Q9X871	Q9x871 streptomyce
56	7	5.3	495	13 Q9W6L0	Q9w6l0 odontaspis
57	7	5.3	496	13 Q9W6L6	Q9w6l6 atropias sup
58	7	5.3	496	13 Q9W6L3	Q9w6l3 lamna ditro
59	7	5.3	497	13 Q9W6L8	Q9w6l8 megachasma
60	7	5.3	497	13 Q9W6K8	Q9w6k8 atropias pel
61	7	5.3	498	13 Q9W6L5	Q9w6l5 iavirus oxyr
62	7	5.3	499	13 Q9W6L7	Q9w6l7 cartharodon
63	7	5.3	499	13 Q9W6L4	Q9w6l4 pseudocarch
64	7	5.3	510	11 Q9CXC9	Q9cxc9 mus musculu
65	7	5.3	510	11 Q9JHB1	Q9jhb1 mus musculu
66	7	5.3	510	11 Q8C0R3	Q8c0r3 mus musculu
67	7	5.3	528	11 Q8BUR1	Q8bur1 mus musculu
68	7	5.3	548	17 Q97VC0	Q97vc0 sulfolobus
69	7	5.3	555	10 Q9L268	Q9l268 arabidopsis
70	7	5.3	567	5 Q8WTC2	Q8wtc2 drosophila
71	7	5.3	622	16 Q8NR65	Q8nr65 corynebacte
72	7	5.3	633	5 Q9VJ41	Q9vja1 drosophila
73	7	5.3	705	2 Q50468	Q50468 mycobacteri
74	7	5.3	705	16 P96283	P96283 mycobacteri
75	7	5.3	720	5 Q8INYL	Q8iny1 drosophila
76	7	5.3	914	4 Q8IVF7	Q8ivf7 homo sapien
77	7	5.3	924	2 Q92IG3	Q92ig3 rhodothermu
78	7	5.3	924	2 Q9X582	Q9x582 rhodothermu
79	7	5.3	1061	16 Q8BDU2	Q8bd02 streptococc
80	7	5.3	1080	5 Q01979	Q01979 caenorhabdi
81	7	5.3	1112	5 Q90381	Q90381 cartharindu
82	7	5.3	1218	5 Q8IU42	Q8iua2 dictyosteli
83	7	5.3	1729	10 Q8LK56	Q8lk56 arabidopsis
84	7	5.3	1765	5 Q8SK63	Q8sk63 drosophila
85	7	5.3	2232	5 Q9VC96	Q9vc96 drosophila
86	7	5.3	2438	5 Q9VQL7	Q9vql7 drosophila
87	7	5.3	2498	13 Q8QG78	Q8qg78 xenopus lae
88	7	5.3	3414	12 Q10383	Q10383 louping ill
89	7	5.3			

90 6 4.6 12 4 016452  
91 6 4.6 22 8 09T2H8  
92 6 4.6 24 8 09BUT7  
93 6 4.6 30 16 08DR90  
94 6 4.6 43 16 08E2E8  
95 6 4.6 48 4 09UE84  
96 6 4.6 51 16 08F1G0  
97 6 4.6 53 2 050092  
98 6 4.6 53 4 060251  
99 6 4.6 54 16 092P27  
100 6 4.6 56 16 08YTE1

## ALIGNMENTS

RESULT 1  
Q9BRT3 PRELIMINARY; PRT; 115 AA.  
ID Q9BRT3  
AC 01-JUN-2001 (TrEMBLrel. 17, Created)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
DE 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
DE Similar to RIKEN CDNA 1810046J19 gene.  
OS Homo sapiens (human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Brain;  
RA Strausberg R.;  
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; BC006006; AAH06006.1; -; 5D8B911C0F23DDC1 CRC64;  
SQ SEQUENCE 115 AA; 12403 MW; 5D8B911C0F23DDC1 CRC64;  
Query Match 87.8%; Score 115; DB 4; Length 115;  
Best Local Similarity 100.0%; Pred. No. 3e-111;  
Matches 115; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 17 MSGEPGQTSVAPPEEVEPGSGVRIIVEYCEPCGFATYIELASAVKEQYPGIEISRLG 76  
Db 1 MSGEPGQTSVAPPEEVEPGSGVRIIVEYCEPCGFATYIELASAVKEQYPGIEISRLG 60  
QY 77 GTGAFETIINGQLVFSKLENGGFPYKDLTFAIRASNGETLEKTTSPPCVTL 131  
Db 61 GTGAFETIINGQLVFSKLENGGFPYKDLTFAIRASNGETLEKTTSPPCVTL 115  
RESULT 2  
Q9CQ86 PRELIMINARY; PRT; 115 AA.  
ID Q9CQ86  
AC 09CQ86;  
DT 01-JUN-2001 (TrEMBLrel. 17, Created)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
DE 1810046J19R1K protein (RIKEN CDNA 1810046J19 gene) (Hypothetical  
DE protein).  
GN 1810046J19R1K.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6J; TISSUE=Brain, Pancreas, and Tongue;  
RX MEDLINE=1085660; PubMed=1121851;  
RA Kawai J., Shimasawa A., Shibata K., Yoshino M., Itoh M., Iehi Y.,  
RA Arikawa T., Hara A., Fukunishi Y., Kono H., Adachi J., Fukuda S.,  
RA Aizawa K., Ozawa M., Nishi K., Kiyosawa H., Kondo S., Yamana K. I.,  
RA Saito T., Okazaki Y., Gojohori T., Bono H., Kasukawa T., Saito R.,  
RA Kadoya K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,  
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochwa H.,

RA Kuenl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,  
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,  
RA Sakai K., Okita T., Furuno M., Aono H., Baldarelli R., Barsh G.,  
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,  
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,  
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamuya M., Lee N.H.,  
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,  
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,  
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,  
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,  
RA Hayaishizaki Y.,  
RT "functional annotation of a full-length mouse CDNA collection."  
RL Nature 409:685-690(2001).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Liver;  
RA Strausberg R.;  
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6J; TISSUE=Aorta and vein;  
RX MEDLINE=22354683; PubMed=1246851;  
RA The FANTOM Consortium,  
RA The RIKEN Genome Exploration Research Group Phase I & II Team,  
RT "Analysis of the mouse transcriptome based on functional annotation of  
RT 60,770 full-length cDNAs."  
RL Nature 420:563-573(2002).  
RL Nature 420:563-573(2002).  
DR EMBL; AK009922; BAB26586.1; -;  
DR EMBL; AK002959; BAB22480.1; -;  
DR EMBL; AK007795; BAB25261.1; -;  
DR EMBL; BC021589; BAB21589.1; -;  
DR EMBL; AK041314; BAC30901.1; -;  
DR MGD; MGI:1913678; 1810046J19R1K.  
KM Hypothetical protein.  
SQ SEQUENCE 115 AA; 12295 MW; B36A0340DFBA737A CRC64;

Query Match 30.5%; Score 40; DB 11; Length 115;  
Best Local Similarity 100.0%; Pred. No. 2.3e-33;  
Matches 40; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 66 YPGIEIESRLGTGAFETIINGQLVFSKLENGGFPYKDL 105  
Db 50 YPGIEIESRLGTGAFETIINGQLVFSKLENGGFPYKDL 89  
RESULT 3  
Q9Y4Y4 PRELIMINARY; PRT; 158 AA.  
ID Q9Y4Y4  
AC 09Y4Y4;  
DT 01-NOV-1999 (TrEMBLrel. 12, Created)  
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
DE B6 protein.  
GN HPV45 B6.  
OS Homo sapiens (human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Sastre-Garau X., Favre M., Couturier J., Orth G.,  
RT "Distinct patterns of alteration of myc genes associated with  
RT integration of human papillomavirus type 16 ce type 45 in two genital  
RT tumours."  
RL J. Gen. Virol. 81:198-199(2000).  
DR EMBL; AJ242956; CAB44706.1; -;  
DR InterPro: IPR001334; B6.  
DR Pfam: PF00518; B6; 1.  
SQ SEQUENCE 158 AA; 18914 MW; A61AEF98390AEBB3 CRC64;  
Query Match 6.9%; Score 9; DB 4; Length 158;  
Best Local Similarity 100.0%; Pred. No. 0.46;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 115 GETLEKITT 123  
 |||||  
 Db 87 GETLEKITT 95

## RESULT 4

010608 PRELIMINARY; PRT; 158 AA.

AC 010608;  
 DT 01-JUL-1997 (T-EMBLrel. 04, Created)  
 DT 01-JUL-1997 (T-EMBLrel. 04, Last sequence update)  
 DE 01-DEC-2001 (T-EMBLrel. 19, Last annotation update)  
 DE Oncoprotein E6.  
 GN E6.  
 OS Human papillomavirus type 45.  
 OC Viruses; dsDNA viruses, no RNA stage; Papillomaviridae;  
 OC Papillomavirus.  
 NCBI\_TaxID=10593;  
 RX SEQUENCE FROM N.A.  
 RP STRAIN=IC4 HPV45 variant;  
 RA Saxtre-Garau X., Favre M., Couturier J., Orth G.;  
 RT "Distinct patterns of alteration of myc genes associated with  
 integration of HPV16 or HPV45 DNA in two genital tumors."  
 RL J. Gen. Virol. 0:0-0(0).  
 DR EMBL; Y13218; CA73660.1; -  
 DR InterPro; IPR001334; E6.  
 DR Pfam; PF00518; E6; 1.  
 SQ SEQUENCE 158 AA; 18914 MW; A61AEF98390AEBB3 CRC64;

Query Match 6.9%; Score 9; DB 12; Length 158;  
 Best Local Similarity 100.0%; Pred. No. 0.46;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 115 GETLEKITT 123  
 |||||  
 Db 87 GETLEKITT 95

## RESULT 5

08S227 PRELIMINARY; PRT; 232 AA.

AC 08S227;  
 DT 01-JUN-2002 (T-EMBLrel. 21, Created)  
 DT 01-JUN-2002 (T-EMBLrel. 21, Last sequence update)  
 DT 01-OCT-2002 (T-EMBLrel. 22, Last annotation update)  
 DE P0446G04.25 protein (P0460C04.2 protein).  
 GN P0446G04.25 OR P0460C04.2.  
 OS Oryza sativa (Japonica cultivar-group).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliopsida; Liliopsida; Poales; Poaceae;  
 OC Ehrhartoideae; Oryzae; Oryza.  
 NCBI\_TaxID=3947;  
 RX SEQUENCE FROM N.A.  
 RP STRAIN=cv. Nipponbare;  
 RA Sasaki T., Matsumoto T., Yamamoto K.;  
 RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC  
 clone:P0446G04.25"  
 RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. Nipponbare;  
 RA Sasaki T., Matsumoto T., Yamamoto K.;  
 RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC  
 clone:P0460C04.25"  
 RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AP003252; BAB89601.1; -  
 DR EMBL; AP004366; BAB92910.1; -  
 DR Gramene; Q8S227; -  
 SQ SEQUENCE 232 AA; 25385 MW; 7011E960E7909E4F CRC64;

Query Match 6.9%; Score 9; DB 10; Length 232;  
 Best Local Similarity 100.0%; Pred. No. 0.65;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 86 NGQVFSKL 94  
 |||||  
 Db 191 NGQVFSKL 199

## RESULT 6

Q50044 PRELIMINARY; PRT; 163 AA.

AC Q50044;  
 DT 01-NOV-1996 (T-EMBLrel. 01, Created)  
 DT 01-NOV-1996 (T-EMBLrel. 01, Last sequence update)  
 DT 01-NOV-1996 (T-EMBLrel. 01, Last annotation update)  
 DE Phox.  
 OS Mycobacterium leprae.  
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
 OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
 NCBI\_TaxID=1769;  
 RX SEQUENCE FROM N.A.  
 RP Smith D.R.;  
 RA Submitted (APR-1995) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Robison K.;  
 RL Submitted (SEP-1994) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; U15182; AAA62989.1; -  
 SQ SEQUENCE 163 AA; 17551 MW; E8F2D0943B18443D CRC64;

Query Match 6.1%; Score 8; DB 2; Length 163;  
 Best Local Similarity 100.0%; Pred. No. 5.2;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 36 GSGVRIVV 43  
 |||||  
 Db 34 GSGVRIVV 41

## RESULT 7

Q9HK29 PRELIMINARY; PRT; 309 AA.

AC Q9HK29;  
 DT 01-MAR-2001 (T-EMBLrel. 16, Created)  
 DT 01-MAR-2001 (T-EMBLrel. 16, Last sequence update)  
 DT 01-OCT-2002 (T-EMBLrel. 22, Last annotation update)  
 DE 2-hydroxyacid dehydrogenase related protein.  
 GN TA0779.  
 OS Thermoplasma acidophilum.  
 OC Archaea; Euryarchaeota; Thermoplasmatata; Thermoplasmatales;  
 OC Thermoplasmataceae; Thermoplasma.  
 NCBI\_TaxID=2303;  
 RX SEQUENCE FROM N.A.  
 RC STRAIN=DSM 1728;  
 RX MEDLINE=20479572; PubMed=11029001;  
 RA Ruepp A., Graml W., Santos-Martinez M.-L., Koretke K.K., Volker C.,  
 RA Mewes H.-W., Frißman D., Stocker S., Lupas A.N., Baumeister W.;  
 RT "The genome sequence of the thermophilic scavenger Thermoplasma  
 acidophilum."  
 RL Nature 407:508-513 (2000).  
 DR EMBL; AL445065; CAC11910.1; -  
 DR HSSP; P36234; 1GDH.  
 DR InterPro; IPR006139; 2-Hacid.DH.  
 DR InterPro; IPR006140; 2-Hacid.DH\_C.  
 DR Pfam; PF00389; 2-Hacid.DH; 1.  
 DR Pfam; PF02826; 2-Hacid.DH\_C; 1.  
 DR PROSITE; PS00670; D\_2\_HYDROXYACID\_DH\_2; 1.  
 DR PROSITE; PS00671; D\_2\_HYDROXYACID\_DH\_3; 1.  
 KW Complete proteome.

SEQ SEQUENCE 309 AA; 34592 MW; 1A2AAD9B0EFD6F5 CRC64;

Query Match 6.1%; Score 8; DB 17; Length 309;  
Best Local Similarity 100.0%; Pred. No. 9.1;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 102 EKDLIEAI 109  
DB 234 EKDLIEAI 241

## RESULT 8

Q8NKN0 PRELIMINARY; PRT; 563 AA.  
AC Q8NKN0;  
DT 01-OCT-2002 (TREMBlrel. 22, Created)  
DT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)  
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)  
DE Putative asparagine synthetase.  
OS uncultured crenarchaeote.  
OC Archaea; Crenarchaeota; environmental samples.  
OX NCBI\_TaxID=29281;

RA "Subversion of host defense mechanisms by adenoviruses."  
RA "First insight into the genome of an uncultivated crenarchaeote from soil."  
RT Environ. Microbiol. 0:0-0(2002).  
RL EMBL; AJ96176; CAD2690.1; -  
DR InterPro: IPR001962; Asn\_synthase.  
DR InterPro: IPR006426; Asn\_synth\_AEB.  
DR InterPro: IPR000583; GATase\_2.  
DR Pfam; PF00733; Asn\_synthase; 1.  
DR TIGRfams; TIGR01536; asn\_synth\_AEB; 1.  
DR PROSITE; PS00443; GATASE\_TYPE\_II; 1.  
SQ SEQUENCE 563 AA; 63302 MW; 76A7BD97B4E298DE CRC64;

Query Match 6.1%; Score 8; DB 1; Length 563;  
Best Local Similarity 100.0%; Pred. No. 15;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 104 DLIEAIRR 111  
DB 133 DLIEAIRR 140

## RESULT 9

Q8UY75 PRELIMINARY; PRT; 106 AA.  
AC Q8UY75;  
DT 01-MAR-2002 (TREMBlrel. 20, Created)  
DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)  
DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)  
DE 11.6 kDa.  
GN E3.  
OS Simian adenovirus 25.  
OC Viruses; dsDNA viruses, no RNA stage; Adenoviridae; Mastadenovirus.  
OX NCBI\_TaxID=175567;

RA "The complete genomic sequence of Mycoplasma penetrans, an intracellular bacterial pathogen in humans."  
RA "Nucleic Acids Res. 30:5293-5300(2002)."  
DR EMBL; AP004172; BAC44363.1; -  
KW Complete proteome.  
SQ SEQUENCE 121 AA; 12450 MW; 97CB6A91F710ADEA CRC64;

Query Match 5.3%; Score 7; DB 16; Length 121;  
Best Local Similarity 100.0%; Pred. No. 44;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

SEQ SEQUENCE 106 AA; 12085 MW; 432EF486B787204C CRC64;

Query Match 5.3%; Score 7; DB 12; Length 106;  
Best Local Similarity 100.0%; Pred. No. 39;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 38 GVRIVE 44  
DB 47 GVRIVE 53

## RESULT 10

Q8BEL7 PRELIMINARY; PRT; 106 AA.  
AC Q8BEL7;  
DT 01-MAR-2003 (TREMBlrel. 23, Created)  
DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)  
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)  
DE E3 12.1 kDa protein.  
OS Human adenovirus type 4.  
OC Viruses; dsDNA viruses, no RNA stage; Adenoviridae; Mastadenovirus.  
OX NCBI\_TaxID=28280;

RA "Subversion of host defense mechanisms by adenoviruses."  
RA "Cur. Top. Microbiol. Immunol. 269:274-319(2002)."  
DR EMBL; AF361223; AAN64738.1; -  
SQ SEQUENCE 106 AA; 12089 MW; 5CFF64A74B1272 CRC64;

Query Match 5.3%; Score 7; DB 12; Length 106;  
Best Local Similarity 100.0%; Pred. No. 39;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 38 GVRIVE 44  
DB 47 GVRIVE 53

## RESULT 11

Q8EVJ1 PRELIMINARY; PRT; 121 AA.  
AC Q8EVJ1;  
DT 01-MAR-2003 (TREMBlrel. 23, Created)  
DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)  
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)  
DE Ribosomal protein L7/L12.  
GN MYP5730.  
OS Mycoplasma penetrans.  
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.  
OX NCBI\_TaxID=28227;

RA "The complete genomic sequence of Mycoplasma penetrans, an intracellular bacterial pathogen in humans."  
RA "Nucleic Acids Res. 30:5293-5300(2002)."  
DR EMBL; AP004172; BAC44363.1; -  
KW Complete proteome.  
SQ SEQUENCE 121 AA; 12450 MW; 97CB6A91F710ADEA CRC64;

Query Match 5.3%; Score 7; DB 16; Length 121;  
Best Local Similarity 100.0%; Pred. No. 44;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 103 KDLEAI 109  
DB 22 KDLEAI 28



## RESULT 12

Q93RX6 PRELIMINARY; PRT; 159 AA.  
 AC O93RX6;  
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
 DE Hypothetical protein SC06296.  
 GN SC06296 OR SCBAC8D1.09C.  
 OS Streptomyces coelicolor.  
 OC Bacteria; Actinobacteria; Actinomycetales;  
 OC Streptomycetales; Streptomycetaceae; Streptomyces.  
 OX NCBI\_TaxID=1902;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=A3(2) / M145;  
 RX MEDLINE=2196410; PubMed=12000953;  
 RA Bentley S.D., Chater K.F., Cerdano-Tarraga A.-M., Challis G.L.,  
 RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kleser H.,  
 RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,  
 RA Huang C.-H., Kleser T., Larke L., Murphy L., Oliver K., O'Neil S.,  
 RA Rabinovitch E., Rajandream M.A., Rutherford K., Rutter S.,  
 RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,  
 RA Warren T., Wierzbicki A., Woodward J., Barrall B.G., Parkhill J.,  
 RA Hopwood D.A.;  
 RT "Complete genome sequence of the model actinomycete Streptomyces  
 coelicolor A3(2).",  
 RL Nature 417:141-147(2002).  
 DR EMBL; AL939127; CAC37898.1; -;  
 DR InterPro; IPR006683; Thioestr\_suf.  
 DR Pfam; PF03061; 4HBT; 1.  
 DR Hypothetical protein; Complete proteome.  
 KW SEQUENCE 159 AA; 17464 MW; 27A094AB2984DA85 CRC64;

Query Match 5.3%; Score 7; DB 16; Length 159;  
 Best Local Similarity 100.0%; Pred. No. 56;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 GPEABPA 15  
 |||||  
 DB 100 GPEABPA 106

## RESULT 13

Q8F207 PRELIMINARY; PRT; 161 AA.  
 AC Q8F207;  
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
 DE Hypothetical protein.  
 GN LA2970.  
 OS Leptospira interrogans.  
 OC Bacteria; Spirochaetes; Spirochaetales; Leptospiraceae; Leptospira.  
 OX NCBI\_TaxID=1173;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=56601 / Serogroup Icterohaemorrhagiae / Serovar 1a1;  
 RA Ren S.;  
 RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AE011460; AAN50169.1; -;  
 DR Hypothetical protein; Complete proteome.  
 KW SEQUENCE 161 AA; 18868 MW; 39D2D687B384F116 CRC64;

Query Match 5.3%; Score 7; DB 16; Length 161;  
 Best Local Similarity 100.0%; Pred. No. 56;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 88 QLVFSKL 94  
 |||||

DB 150 QLVFSKL 156

## RESULT 14

Q8D6L4 PRELIMINARY; PRT; 202 AA.  
 AC Q8D6L4;  
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
 DE Phosphomannose isomerase.  
 GN VV20515.  
 OS Vibrio vulnificus.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;  
 OC Vibrionaceae; Vibrio.  
 OX NCBI\_TaxID=672;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CMCP6;  
 RA Rhee J.H., Kim S.Y., Chung S.S., Kim J.T., Moon Y.H., Jeong H.,  
 RA Choy H.E.;  
 RT "Complete genome sequence of Vibrio vulnificus CMCP6.",  
 RL Submitted (DEC-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AE016809; AA007465.1; -;  
 KW Isomerase; Complete proteome.  
 KW SEQUENCE 202 AA; 22031 MW; 0D27434A388FE635 CRC64;

Query Match 5.3%; Score 7; DB 16; Length 202;  
 Best Local Similarity 100.0%; Pred. No. 68;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 112 ASNGETL 118  
 |||||  
 DB 165 ASNGETL 171

## RESULT 15

Q8ESN0 PRELIMINARY; PRT; 207 AA.  
 AC Q8ESN0;  
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
 DE Hypothetical protein.  
 GN OB0596.  
 OS Oceanobacillus thelyensis.  
 OC Bacteria; Firmicutes; Bacillales; Oceanobacillus.  
 OX NCBI\_TaxID=182710;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=HTB31 / DSM 14371 / JCM 11309;  
 RX MEDLINE=22220767; PubMed=12235376;  
 RA Takami H., Takaki Y., Uchiyama T.;  
 RT "Genome sequence of Oceanobacillus thelyensis isolated from the Iheya  
 RT Ridge and its unexpected adaptive capabilities to extreme  
 RT environments.",  
 RL Nucleic Acids Res. 30:3927-3935(2002).  
 DR EMBL; AP004595; BAC12552.1; -;  
 DR Hypothetical protein; Complete proteome.  
 KW SEQUENCE 207 AA; 23037 MW; 1BFE3648A53B33C8 CRC64;

Query Match 5.3%; Score 7; DB 16; Length 207;  
 Best Local Similarity 100.0%; Pred. No. 70;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 62 VKEQYPG 68  
 |||||  
 DB 42 VKEQYPG 48

## RESULT 16

Q8VE68 PRELIMINARY; PRT; 209 AA.  
 ID Q8VE68

```

AC 08VE68;
DT 01-MAR-2002 (TEMBLrel. 20, Created)
DT 01-MAR-2002 (TEMBLrel. 20, Last sequence update)
DT 01-OCT-2002 (TEMBLrel. 22, Last annotation update)
DE RIKEN CDNA 2310016C16 gene.
GN 2310016C16R1K.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC019664; AAH19664.1; -
DR MGD; MGI:1916840; 2310016C16R1K.
DR InterPro; IPR000889; Glut_peroxidase.
DR Pfam; PF00255; GSHpx; 1.
DR PRINTS; PRO1011; GLUTPROXDAF.
DR PROSITE; PS00763; GLUTATHIONE_PEROXID_2; 1.
SQ SEQUENCE 209 AA; 24160 MW; 8DASCEC1B4ECF3C CRC64;

Query Match 5.3%; Score 7; DB 11; Length 209;
Best Local Similarity 100.0%; Pred. No. 71;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 GPEAEP 15
DB 144 GPEAEP 150

RESULT 17
Q9D7B7 PRELIMINARY; PRT; 209 AA.
AC Q9D7B7;
DT 01-JUN-2001 (TEMBLrel. 17, Created)
DT 01-JUN-2001 (TEMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TEMBLrel. 19, Last annotation update)
DE 2310016C16R1K protein.
GN 2310016C16R1K.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=C57BL/6J; TISSUE=Tongue;
RX MEDLINE=21085660; Pubmed=11217851;
RA Kawai J.; Shinagawa A.; Shibata K.; Yoshino M.; Itoh M.; Ishii Y.;
RA Atakawa T.; Hara A.; Fukunishi Y.; Komo H.; Adachi J.; Fukuda S.;
RA Aizawa K.; Izawa M.; Nishi K.; Kiyosawa H.; Kondo S.; Yamataka I.;
RA Saito T.; Okazaki Y.; Gojohori T.; Bono H.; Kasukawa T.; Saito R.;
RA Kadoya K.; Matsuda H.A.; Ashburner M.; Batalov S.; Casavant T.;
RA Fleischmann W.; Gaasterland T.; Gissi C.; King B.; Kochiwa H.;
RA Kuehl P.; Lewis S.; Matsuo Y.; Nikaide I.; Pesole G.; Quackenbush J.;
RA Schirral L.M.; Staudli F.; Suzuki R.; Tomita M.; Wagner L.; Maehio T.;
RA Sakai K.; Okido T.; Furuno M.; Aono H.; Baldarelli R.; Barin G.;
RA Blake J.; Boffelli D.; Bojunga N.; Carninci P.; de Bonaldo M.F.;
RA Brownstein M.J.; Bult C.; Fletcher C.; Fujita M.; Gariboldi M.;
RA Gustincich S.; Hill D.; Hofmann M.; Hume D.A.; Kantiya M.; Lee N.H.;
RA Lyons P.; Marchionni L.; Mashima J.; Mazzarelli J.; Mombaerts P.;
RA Nordone P.; Ring B.; Ringwald M.; Rodriguez I.; Sakamoto N.;
RA Sasaki H.; Sato K.; Schoenbach C.; Seta T.; Shibata Y.; Storch K.-F.;
RA Suzuki H.; Toyooka K.; Wang K.H.; Weitz C.; Whitaker C.; Wilting L.;
RA Wyszewski A.; Yoshida K.; Hasegawa Y.; Kawaji H.; Kohsaki S.;
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DR EMBL; AK009378; BAB26254.1; -
DR HSSP; P00435; 1GP1.
DR MGD; MGI:1916840; 2310016C16R1K.
DR InterPro; IPR000889; Glut_peroxidase.
DR Pfam; PF00255; GSHpx; 1.

```

```

DR PRINTS; PRO1011; GLUTPROXDAF.
DR PROSITE; PS00763; GLUTATHIONE_PEROXID_2; 1.
SQ SEQUENCE 209 AA; 24148 MW; 8DA93ED1B4ECF21 CRC64;

Query Match 5.3%; Score 7; DB 11; Length 209;
Best Local Similarity 100.0%; Pred. No. 71;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 GPEAEP 15
DB 144 GPEAEP 150

RESULT 18
Q8AOP1 PRELIMINARY; PRT; 212 AA.
AC Q8AOP1;
DT 01-MAR-2003 (TEMBLrel. 23, Created)
DT 01-MAR-2003 (TEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TEMBLrel. 23, Last annotation update)
DE Nef protein (fragment).
GN NEF.
OS Human immunodeficiency virus 1.
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX NCBI_TaxID=11676;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=NYUW-84;
RA Chakraborty R.; Reins M.; John M.; Philpott S.; Rostrom T.;
RA Yang H.B.; Palakudy T. Sr.; Musoke R.; D'Agostino A.; Webster B.;
RA Burger H.; Rowland-Jones S.L.;
RT "Nef characterization and survival among HIV-1 infected African
RT children: evidence of CTL escape.";
RL Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF538558; AAH40109.1; -
SQ SEQUENCE 212 AA; 24281 MW; D756DC5E3CC4B240 CRC64;

Query Match 5.3%; Score 7; DB 15; Length 212;
Best Local Similarity 100.0%; Pred. No. 71;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 PEAEP 16
DB 25 PEAEP 31

RESULT 19
Q8ZJZ0 PRELIMINARY; PRT; 233 AA.
AC Q8ZJZ0;
DT 01-MAR-2002 (TEMBLrel. 20, Created)
DT 01-MAR-2002 (TEMBLrel. 20, Last sequence update)
DT 01-OCT-2002 (TEMBLrel. 22, Last annotation update)
DE Putative inner membrane protein.
GN STM4522.
OS Salmonella typhimurium.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=602;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=LT2 / SGSC1412 / ATCC 700720;
RX MEDLINE=21534948; Pubmed=11677609;
RA McClelland M.; Sanderson K.E.; Spieth J.; Clifton S.W.; Latreille P.;
RA Courtney L.; Porwollik S.; Ali J.; Dante M.; Du F.; Hou S.; Layman D.;
RA Leonard S.; Nguyen C.; Scott K.; Holmes A.; Grewal N.; Mulvaney E.;
RA Ryan E.; Sun H.; Flores L.; Miller W.; Stoneking T.; Nhan M.;
RA Waterston R.; Wilson R.K.;
RT "Complete genome sequence of Salmonella enterica serovar Typhimurium
RT LT2.";
RL Nature 413:852-856(2001).
DR EMBL; AE008913; AAL23340.1; -

```

KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 233 AA; 26998 MW; A9040DD1CD8A19A CRC64;

Query Match 5.3%; Score 7; DB 16; Length 233;  
Best Local Similarity 100.0%; Pred. No. 78;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 118 LEKITS 124  
DB 81 LEKITS 87

RESULT 20

Q8VWA4 PRELIMINARY; PRT; 243 AA.

AC Q8VWA4; 01-MAR-2002 (TEMBLrel. 20, Created)  
DT 01-MAR-2002 (TEMBLrel. 20, Last sequence update)  
DE 01-JUN-2002 (TEMBLrel. 21, Last annotation update)

GN ACIP.

OS Streptomyces galliaenus.  
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
OC Streptomyces; Streptomycetaceae; Streptomyces.

OX NCBI\_TaxID=33899;

RP SEQUENCE FROM N.A.

RC STRAIN=3AR-33;  
RA Chung J., Fujii I., Tsukamoto N., Sankawa U., Ebizuka Y.;  
RT "Axi-vivone-actinomyces biosynthesis gene cluster from Streptomyces  
RT galliaenus."

RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AB008466; BAB72055.1; -  
DR InterPro; IPR000051; SAM bind.

SQ SEQUENCE 243 AA; 26398 MW; C33CF0AF0B5A04A6 CRC64;

Query Match 5.3%; Score 7; DB 2; Length 243;  
Best Local Similarity 100.0%; Pred. No. 80;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 34 EPSSGVR 40  
DB 186 EPSSGVR 192

RESULT 21

Q9NSY0 PRELIMINARY; PRT; 258 AA.

AC Q9NSY0; 01-OCT-2000 (TEMBLrel. 15, Created)  
DT 01-OCT-2000 (TEMBLrel. 15, Last sequence update)  
DE 01-MAR-2003 (TEMBLrel. 23, Last annotation update)

GN DKFZP434P086.

OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

OX NCBI\_TaxID=9606;

RP SEQUENCE FROM N.A.

RC TISSUE=Testis;  
RA Koehler K., Beyer A., Mewes H.W., Gassenhber J., Wiemann S.;  
RT Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.

DR EMBL; AL137662; CAB70864.1; -  
DR InterPro; IPR000719; PTC\_Kinase.

DR Pfam; PF00069; PKinase; 1.  
DR ProDom; PD000001; PTC\_Kinase; 1.

KW Hypothetical protein; ATP-binding; Transferase.  
SQ SEQUENCE 258 AA; 29852 MW; 3C886AD4CDE26EA CRC64;

Query Match 5.3%; Score 7; DB 4; Length 258;  
Best Local Similarity 100.0%; Pred. No. 85;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 27 APPREEV 33  
DB 155 APPREEV 161

RESULT 22

Q34774 PRELIMINARY; PRT; 280 AA.

AC Q34774; 01-JAN-1998 (TEMBLrel. 05, Created)  
DT 01-JAN-1998 (TEMBLrel. 05, Last sequence update)  
DE 01-MAR-2002 (TEMBLrel. 20, Last annotation update)

GN YOB.

OS Bacillus subtilis.  
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.

OX NCBI\_TaxID=1423;

RP SEQUENCE FROM N.A.  
RA Lapidus A., Galleron N., Sorokin A., Ehrlich D.;  
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.

RN [2]

RP SEQUENCE FROM N.A.  
RC STRAIN=168;  
RX MEDLINE=98044033; PubMed=9384377;  
RA Kunze F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,  
RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,  
RA Borriss R., Bourstier L., Brans A., Braun M., Brignell S.C., Bron S.,  
RA Broutelle S., Brusch C.V., Caldwell B., Capuano V., Carter N.M.,  
RA Chol S.K., Codani J.J., Conneron I.F., Cummings N.J., Daniel R.A.,  
RA Denisot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,  
RA Ertter K.D., Errington J., Fabret C., Ferrati E., Foulger D.,  
RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,  
RA Chim S.Y., Glaeser P., Goffeau A., Gollighly E.J., Grandi G.,  
RA Gutseppl G., Guy B.J., Haga K., Haeck U., Harwood C.R., Henaut A.,  
RA Hilbert H., Holsappel S., Hosono S., Hult M.F., Itaya M., Jones L.,  
RA Joris B., Karamata D., Kasahara Y., Klaerner-Blanchard M., Klein C.,  
RA Kobayashi Y., Koelter P., Koningsstein G., Krogh S., Kumano M.,  
RA Kurita K., Lapidus A., Lardinois S., Lauder J., Lazarevic V.,  
RA Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,  
RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,  
RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega H., Park S.H.,  
RA Parro V., Pohl T.M., Portetlelle D., Portollik S., Prescott A.M.,  
RA Presecan E., Pujić P., Purnelle B., Rapoport G., Rey M., Reynolds S.,  
RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadale Y.,  
RA Sato T., Scanlan B., Schleich S., Schroeder R., Scofield F.,  
RA Sekiguchi J., Sekowska A., Seror S.J., Seror P., Shin B.S., Soldo B.,  
RA Sorokin A., Taccioni E., Takagi T., Takahashi H., Takemaru K.,  
RA Takeuchi M., Tamakoshi A., Tanaka T., Terpsira P., Tognoni A.,  
RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassaretto A.,  
RA Viari A., Wandt R., Wedler E., Wedler H., Weitzenecker T.,  
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasuno K., Yata K.,  
RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.,  
RT "The complete genome sequence of the gram-positive bacterium Bacillus  
RT subtilis."

RL Nature 390:249-256 (1997).

RN [3]

RP SEQUENCE FROM N.A.  
RC STRAIN=168;  
RA Kunze F., Ogasawara N., Yoshikawa H., Danchin A.;  
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.

DR EMBL; AF027868; AAB84462.1; -  
DR EMBL; Z59114; CAB13790.1; -

KW Complete proteome.  
SQ SEQUENCE 280 AA; 33749 MW; 3E8921D042F8068 CRC64;

Query Match 5.3%; Score 7; DB 16; Length 280;  
Best Local Similarity 100.0%; Pred. No. 91;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 43 GETLEKI 49

## RESULT 23

ID Q9RUX4 PRELIMINARY; PRT; 287 AA.

AC Q9RUX4;

DT 01-MAY-2000 (TrEMBLrel. 13, Created)

DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)

DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)

DE Hypothetical protein SC01196.

GN SC01196 OR SCG11A.27C.

OS Streptomyces coelicolor.

OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales; Streptomycetales; Streptomycetaceae; Streptomyces.

NCBI\_TaxID=1902;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=AJ(2) / M145;

RX MEDLINE=21996410; PubMed=12000953;

RA Bentley S.D., Chater K.F., Cerdano-Parraga A.-M., Challis G.L.,

RA Thompson N.R., James K.D., Harris D.E., Quail M.A., Kleser H.,

RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,

RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,

RA Huang C.-H., Kleser T., Larke L., Murphy L., Oliver K., O'Neill S.,

RA Rabinowitch E., Rajadream M.A., Rutherford K., Rutter S.,

RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,

RA Warren T., Wierczek A., Woodward J., Barrell B.G., Parkhill J.,

RA Hopwood D.A.;

RT "Complete genome sequence of the model actinomycete Streptomyces

RT coelicolor AJ(2).";

RL Nature 417:141-147(2002).

DR EMBL: AL39108; CAB61608.1; -.

DR InterPro: IPR006311; Tat.

DR TIGRfams: TIGR01409; TAT\_signal; seq; 1.

DR Hypothetical protein; Complete proteome.

SQ SEQUENCE 287 AA; 31470 MW; E6DA80924502C3 CRC64;

Query Match 5.3%; Score 7; DB 16; Length 287;

Best Local Similarity 100.0%; Pred. No. 93;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 52 EATYLEL 58

Db 130 EATYLEL 136

## RESULT 24

Q8WYS5

ID Q8WYS5 PRELIMINARY; PRT; 293 AA.

AC Q8WYS5;

DT 01-MAR-2002 (TrEMBLrel. 20, Created)

DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)

DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)

DE Hypothetical protein.

GN PP9320.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

NCBI\_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RA Zhou X.M., Zhang P.P., Jiang H.Q., Huang Y., Qin W.X., Zhao X.T.,

RA Man D.F., Gu J.R.;

RT "Novel human cDNA clones with function of inhibiting cancer cell

RT growth.";

RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.

DR EMBL: AF318376; AAL55883.1; -.

DR InterPro: IPR000719; Prot\_kinase.

DR Pfam: PF00069; Pkinase\_1.

DR Hypothetical protein; ATP-binding; Transferase.

SQ SEQUENCE 293 AA; 33141 MW; B37F0F45BD8AC99B CRC64;

Query Match 5.3%; Score 7; DB 4; Length 293;

Best Local Similarity 100.0%; Pred. No. 95;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 27 APPPEEV 33

Db 190 APPPEEV 196

## RESULT 25

Q8NCX8

ID Q8NCX8 PRELIMINARY; PRT; 294 AA.

AC Q8NCX8;

DT 01-OCT-2002 (TrEMBLrel. 22, Created)

DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)

DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)

DE Hypothetical protein (Fragment).

GN DKFP43412411.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

NCBI\_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Testis;

RA Diesterhoef A., Lauber J., Mewes H.W., Weil B., Wiemann S.;

RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.

DR EMBL: AL834530; CAD39186.1; -.

DR InterPro: IPR000719; Prot\_kinase.

DR Pfam: PF00069; Pkinase\_1.

DR ProDom: PD000001; Prot\_kinase; 1.

DR PROSITE: PS50011; PROTEIN\_KINASE\_DOM; 1.

DR Hypothetical protein; ATP-binding; Transferase.

SQ SEQUENCE 294 AA; 33958 MW; D0FF57F1901AFB18 CRC64;

Query Match 5.3%; Score 7; DB 4; Length 294;

Best Local Similarity 100.0%; Pred. No. 95;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 27 APPPEEV 33

Db 191 APPPEEV 197

## RESULT 26

Q9CL61

ID Q9CL61 PRELIMINARY; PRT; 306 AA.

AC Q9CL61;

DT 01-JUN-2001 (TrEMBLrel. 17, Created)

DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)

DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)

DE SpeE.

GN SpeE OR PM1381.

OS Pasteurella multocida.

OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;

OC Pasteurellaceae; Pasteurella.

NCBI\_TaxID=747;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=Pm70;

RA May B.J., Zhang O., Li L.L., Paustian M.L., Whitam T.S., Kapur V.;

RT "Complete genomic sequence of Pasteurella multocida Pm70.";

RL Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465(2001).

DR EMBL: AE006176; AAK03465.1; -.

DR HSP: P53608; 2CEV.

DR InterPro: IPR005925; Agmatinase.

DR Pfam: PF00491; arginase\_1.

DR TIGRfams: TIGR01230; agmatinase; 1.

DR PROSITE: PS00147; ARGINASE\_1; 1.

DR PROSITE: PS00148; ARGINASE\_2; 1.

DR PROSITE; PS01053; ARGINASE\_3, 1.  
 KW Complete proteome.  
 SQ SEQUENCE 306 AA; 33708 MW; 355D609504F4EC3B CRC64;

Query Match 5.3%; Score 7; DB 16; Length 306;  
 Best Local Similarity 100.0%; Pred. No. 98;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 107 EAIRRAS 113  
 DB 58 EAIRRAS 64

RESULT 27

09JRG2 PRELIMINARY; PRT; 307 AA.

AC 09JRG2;  
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE Agmatinase (EC 3.5.3.11).  
 GN SPEB OR NMA2016 OR NMB0469.  
 OS Neisseria meningitidis (serogroup A), and  
 OS Neisseria meningitidis (serogroup B).  
 OC Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;  
 OC Neisseriaceae; Neisseria.  
 NCBI\_TaxID=65699, 491;  
 [1] \_

SEQUENCE FROM N.A.

RC STRAIN=22491 / Serogroup A / Serotype 4A;  
 RX MEDLINE=20222556; PubMed=10761919;  
 RA Parthill J., Achtman M., James K.D., Bentley S.D., Churcher C.,  
 Klee S.R., Morelli G., Basham D., Brown D., Chillingworth T.,  
 Davies R.M., Davis P., Devlin K., Felwell T., Hamlin N., Holroyd S.,  
 Jagals K., Leather S., Moule S., Mungall K., Quail M.A.,  
 Rajandream M.A., Rutherford K.M., Simmonds M., Skelton J.,  
 Whitehead S., Spratt B.G., Barrall B.G.;  
 RT "Complete DNA sequence of a serogroup A strain of Neisseria  
 meningitidis 22491."  
 RL Nature 404:502-506 (2000).  
 RN [2]

SEQUENCE FROM N.A.

RC STRAIN=MC58 / Serogroup B;  
 RX MEDLINE=20175755; PubMed=10710307;  
 RA Tettelin H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E.,  
 Eisen J.A., Ketchum K.A., Hood D.W., Peden J.F., Dodson R.J.,  
 Nelson W.C., Gwin M.L., Deboy R., Peterson J.D., Hickey E.K.,  
 Haft D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A.,  
 Mason T., Clecko A., Parksey D.S., Blair E., Ciftone H., Clark E.B.,  
 Cotton M.D., Utterback T.R., Khouri H., Qin H., Vamathevan J.,  
 Gill J., Scarlato V., Maignani V., Pizzo M., Grandi G., Sun L.,  
 Smith H.O., Fraser C.M., Moxon E.R., Rappelli R., Venter J.C.;  
 RT "Complete genome sequence of Neisseria meningitidis serogroup B strain  
 MC58."  
 RL Science 287:1809-1815 (2000).  
 DR EMBL; AL162757; CAB85235.1; -;  
 DR EMBL; AE002403; AAF40906.1; -;  
 DR HSSP; P53608; 2CEV.  
 DR TIGR; NMB0469; -;  
 DR InterPro; IPR005925; Agmatinase.  
 DR InterPro; IPR006035; Arg\_agn\_form.  
 DR InterPro; IPR000834; Zn\_carboxypept.  
 DR Pfam; PF00491; arginase; 1.  
 DR TIGRPFAM; TIGR01230; agmatinase; 1.  
 DR PROSITE; PS00147; ARGINASE\_1; 1.  
 DR PROSITE; PS00148; ARGINASE\_2; 1.  
 DR PROSITE; PS01053; ARGINASE\_3; 1.  
 DR PROSITE; PS00133; CARBOXYPEPT\_ZN\_2; 1.  
 DR Hydrolase; Complete proteome.  
 SQ SEQUENCE 307 AA; 33905 MW; 1B5013800B47E4BF CRC64;

Query Match 5.3%; Score 7; DB 16; Length 307;  
 Best Local Similarity 100.0%; Pred. No. 99;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 107 EAIRRAS 113  
 DB 59 EAIRRAS 65

RESULT 28

09ZWT8 PRELIMINARY; PRT; 319 AA.

AC 09ZWT8;  
 DT 01-MAY-1999 (TrEMBLrel. 10, Created)  
 DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)  
 DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)  
 DE Putrescine N-methyltransferase (Fragmant).  
 GN NSPMT2.  
 OS Nicotiana sylvestris (Wood tobacco).  
 OS Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 OC Asteridae; lamids; Solanales; Solanaceae; Nicotiana.  
 NCBI\_TaxID=4096;  
 RN [1]

SEQUENCE FROM N.A.

RX MEDLINE=98281572; PubMed=9620262;  
 RA Hashimoto T., Shoji T., Mihara T., Oguri H., Tamaki K., Suzuki K.,  
 Yamada Y.;  
 RT "Intraspecific variability of the tandem repeats in Nicotiana  
 putrescine N-methyltransferase."  
 RL Plant Mol. Biol. 37:25-37 (1998).  
 RN [2]

SEQUENCE FROM N.A.

RX MEDLINE=20419123; PubMed=10965939;  
 RA Shoji T., Yamada Y., Hashimoto T.;  
 RT "Amonate induction of putrescine N-methyltransferase genes in the  
 root of Nicotiana sylvestris."  
 RL Plant Cell Physiol. 41:831-839 (2000).  
 DR EMBL; AB04323; BAA74543.1; -;  
 DR InterPro; IPR000051; SAM\_bind.  
 DR InterPro; IPR001045; Spermine synthase.  
 DR Pfam; PF01564; Spermine synth. 1.  
 DR PROSITE; PS01330; SPERMIDINE\_SYNTHASE; 1.  
 KW Methyltransferase; Transferase.  
 FT NON TER 319 319  
 SQ SEQUENCE 319 AA; 34938 MW; 864A4DB0E0728502 CRC64;

Query Match 5.3%; Score 7; DB 10; Length 319;  
 Best Local Similarity 100.0%; Pred. No. 1e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 95 ENGCFPY 101  
 DB 147 ENGCFPY 153

RESULT 29

08TYL1 PRELIMINARY; PRT; 322 AA.

AC 08TYL1;  
 DT 01-JUN-2002 (TrEMBLrel. 21, Created)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
 DE Uncharacterized protein conserved in archaea.  
 GN MK0285.

Methanopyrus kandleri.

OC Archaea; Euryarchaeota; Methanopyri; Methanopyrales; Methanopyraceae;  
 OC Methanopyrus.

NCBI\_TaxID=2320;  
 RN [1]

SEQUENCE FROM N.A.

RC STRAIN=AV19 / DSM 6324 / JCM 9639;  
 RX MEDLINE=21927647; PubMed=11930014;  
 RA Slesarev A.I., Mezhevaya K.V., Makarova K.S., Polushin N.N.,  
 Shcherbinina O.V., Shakhova V.V., Belova G.I., Aravind L.,

RA Natale D.A., Rogozin I.B., Tatusov R.L., Wolf Y.I., Stetter K.O.,  
 RA Malykh A.G., Koonin E.V., Kozayavkin S.A.,  
 RT "The complete genome of hyperthermophile Methanopyrus kandleri AV19  
 RT and monophyly of archaeal methanogens.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:4644-4649 (2002).  
 DR EMBL; AE010326; AAM01502.1; -.  
 KW Complete proteome.  
 SQ SEQUENCE 322 AA; 35899 MW; 2AIDP166797A01E3 CRC64;

Query Match 5.3%; Score 7; DB 17; Length 322;  
 Best Local Similarity 100.0%; Pred. No. 1.e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 30 PERVEPG 36  
 DB 286 PERVEPG 292

RESULT 30  
 ID Q9S7W8 PRELIMINARY; PRT; 336 AA.  
 AC Q9S7W8;  
 DT 01-MAY-2000 (TREMBlrel. 13, Created)  
 DT 01-MAY-2000 (TREMBlrel. 13, last sequence update)  
 DT 01-JUN-2002 (TREMBlrel. 21, last annotation update)  
 DE Putrescine N-methyltransferase 1 (EC 2.1.1.53).  
 GN ABPM1.  
 OS Atropa belladonna (Belladonna) (Deadly nightshade).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 OC Asteridae; lamids; Solanales; Solanaceae; Atropa.  
 NCBI\_TaxID=33113;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=99281483; PubMed=10353217;  
 RA Suzuki K., Yamada Y., Hashimoto T.;  
 RT "Expression of Atropa belladonna putrescine N-methyltransferase gene in  
 RT root pericycle.";  
 RL Plant Cell Physiol. 40:289-297 (1999).  
 DR EMBL; AB018573; BAA82264.1; -.  
 DR EMBL; AB018570; BAA82261.1; -.  
 DR InterPro: IPR000051; SAM bind.  
 DR InterPro: IPR001045; Spmrine synthase.  
 DR Pfam: PF01564; Spermine\_synth. 1.  
 DR TrGFAMs: TIGR00417; spe. 1.  
 DR PROSITE: PS01330; SPERMIDINE\_SYNTHASE; 1.  
 DR KMW MetHyltransferase; Transferase.  
 SQ SEQUENCE 336 AA; 36976 MW; 30BB9CD87E04AF29 CRC64;

Query Match 5.3%; Score 7; DB 10; Length 336;  
 Best Local Similarity 100.0%; Pred. No. 1.e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 95 ENGPFY 101  
 DB 103 ENGPFY 109

RESULT 31  
 ID Q9W6K9 PRELIMINARY; PRT; 336 AA.  
 AC Q9W6K9;  
 DT 01-NOV-1999 (TREMBlrel. 12, Created)  
 DT 01-NOV-1999 (TREMBlrel. 12, last sequence update)  
 DT 01-DEC-2001 (TREMBlrel. 19, last annotation update)  
 DE Recombination-activating protein RAG-1 (Fragment).  
 GN RAG-1.  
 OS Odonaspis ferox.  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Chondrichthyes;  
 OC Elasmobranchii; Galeomorphii; Galeoidea; Lamniformes; Odonaspididae;  
 OC Odonaspis.  
 NCBI\_TaxID=57989;  
 RN [1]

RP SEQUENCE FROM N.A.  
 RX MEDLINE=99334619; PubMed=10406116;  
 RA Martin A.P.;  
 RT "Substitution rates of organelle and nuclear genes in sharks;  
 RT implicating metabolic rate.";  
 RL Mol. Biol. Evol. 16:996-1002 (1999).  
 DR EMBL; AF135474; AAD32633.1; -.  
 FT NON\_TER 1  
 FT NON\_TER 1  
 SQ SEQUENCE 336 AA; 37798 MW; 00AA5E682A0D948A CRC64;

Query Match 5.3%; Score 7; DB 13; Length 336;  
 Best Local Similarity 100.0%; Pred. No. 1.e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 58 LASAVKE 64  
 DB 82 LASAVKE 88

RESULT 32  
 ID Q9W6L1 PRELIMINARY; PRT; 336 AA.  
 AC Q9W6L1;  
 DT 01-NOV-1999 (TREMBlrel. 12, Created)  
 DT 01-NOV-1999 (TREMBlrel. 12, last sequence update)  
 DT 01-DEC-2001 (TREMBlrel. 19, last annotation update)  
 DE Recombination-activating protein RAG-1 (Fragment).  
 GN RAG-1.  
 OS Crotaphytus maximus.  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Chondrichthyes;  
 OC Elasmobranchii; Galeomorphii; Galeoidea; Lamniformes; Alopiidae;  
 OC Cetorhinus.  
 NCBI\_TaxID=57982;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=99334619; PubMed=10406116;  
 RA Martin A.P.;  
 RT "Substitution rates of organelle and nuclear genes in sharks;  
 RT implicating metabolic rate.";  
 RL Mol. Biol. Evol. 16:996-1002 (1999).  
 DR EMBL; AF135476; AAD32635.1; -.  
 FT NON\_TER 1  
 FT NON\_TER 1  
 SQ SEQUENCE 336 AA; 37728 MW; 39077115B28CC322 CRC64;

Query Match 5.3%; Score 7; DB 13; Length 336;  
 Best Local Similarity 100.0%; Pred. No. 1.e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 58 LASAVKE 64  
 DB 82 LASAVKE 88

RESULT 33  
 ID Q8XRS2 PRELIMINARY; PRT; 337 AA.  
 AC Q8XRS2;  
 DT 01-MAR-2002 (TREMBlrel. 20, Created)  
 DT 01-MAR-2002 (TREMBlrel. 20, last sequence update)  
 DT 01-MAR-2002 (TREMBlrel. 20, last annotation update)  
 DE Hypothetical protein Rsp0759.  
 GN RSP0759 OR RSP01949.  
 OS Ralstonia solanacearum (Pseudomonas solanacearum).  
 OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;  
 OC Ralstoniaceae; Ralstonia.  
 NCBI\_TaxID=305;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=GMI1000;  
 RX MEDLINE=21681879; PubMed=11823852;

RA Salanoubat M., Genin S., Artiguenave F., Guzy J., Mangenot S.,  
 RA Arlat M., Billault A., Brothier P., Camus J.C., Catolico L.,  
 RA Chandel M., Choise N., Claudel-Renard C., Cunac S., Demange N.,  
 RA Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schiex T.,  
 RA Sigulier P., Thebaud P., Whalen M., Wincker P., Levy M.,  
 RA Weisenbach J., Boucher C.A.,  
 RT "genome sequence of the plant pathogen Ralstonia solanacearum";  
 RL Nature 415:497-502(2002).  
 DR EMBL; AF46080; CAD17910.1; -;  
 KM Plasmid; Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 337 AA; 37094 MW; CIEE68258B839B97 CRC64;

Query Match 5.3%; Score 7; DB 16; Length 337;  
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 105 LIBAIR 111  
 Db 141 LIBAIR 147

## RESULT 34

Q9XJ41 PRELIMINARY; PRT; 338 AA.  
 AC Q9XJ41;  
 DT 01-NOV-1999 (TREMBlrel. 12, Created)  
 DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
 DE Putrescine N-methyltransferase (EC 2.1.1.53).  
 GN HNPMT.  
 OS Hyoscyamus niger (Henbane).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 OC Asteridae; lamids; Solanales; Solanaceae; Hyoscyamus.  
 OX NCBI\_TaxID=4079;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Root;  
 RX MEDLINE=99281483; PubMed=10353217;  
 RA Suzuki K., Yamada Y., Hashimoto T.;  
 RT "Expression of Atropa belladonna putrescine N-methyltransferase gene in  
 RT root pericycle.";  
 RL Plant Cell Physiol. 40:289-297(1999).  
 DR EMBL; AB018572; BAA82262.1; -;  
 DR InterPro; IPR000051; SAM\_bind.  
 DR InterPro; IPR001045; Sprime\_synthase.  
 DR Pfam; PF01564; Spermine\_synth; 1.  
 DR TIGRPFAM; TIGR00417; spse; 1.  
 DR PROSITE; PS01330; SPERMIDINE\_SYNTHASE; 1.  
 DR Methyitransferase; Transferase.  
 SQ SEQUENCE 338 AA; 37202 MW; A1E97AFPB29FA63 CRC64;

Query Match 5.3%; Score 7; DB 10; Length 338;  
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 95 ENGFPY 101  
 Db 105 ENGFPY 111

## RESULT 35

Q9XJ42 PRELIMINARY; PRT; 340 AA.  
 AC Q9XJ42;  
 DT 01-NOV-1999 (TREMBlrel. 12, Created)  
 DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
 DE Putrescine N-methyltransferase 2 (EC 2.1.1.53).  
 GN ABPMT.  
 OS Atropa belladonna (Belladonna) (Deadly nightshade).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

OC Asteridae; lamids; Solanales; Solanaceae; Atropa.  
 OX NCBI\_TaxID=33113;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Root;  
 RX MEDLINE=99281483; PubMed=10353217;  
 RA Suzuki K., Yamada Y., Hashimoto T.;  
 RT "Expression of Atropa belladonna putrescine N-methyltransferase gene in  
 RT root pericycle.";  
 RL Plant Cell Physiol. 40:289-297(1999).  
 DR EMBL; AB018572; BAA82262.1; -;  
 DR InterPro; IPR000051; SAM\_bind.  
 DR InterPro; IPR001045; Sprime\_synthase.  
 DR Pfam; PF01564; Spermine\_synth; 1.  
 DR TIGRPFAM; TIGR00417; spse; 1.  
 DR PROSITE; PS01330; SPERMIDINE\_SYNTHASE; 1.  
 KM Methyitransferase; Transferase.  
 SQ SEQUENCE 340 AA; 37463 MW; 6CC5C0DA012B0A64 CRC64;

Query Match 5.3%; Score 7; DB 10; Length 340;  
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 95 ENGFPY 101  
 Db 104 ENGFPY 110

## RESULT 36

Q9ZWT9 PRELIMINARY; PRT; 353 AA.  
 AC Q9ZWT9;  
 DT 01-MAY-1999 (TREMBlrel. 10, Created)  
 DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
 DE Putrescine N-methyltransferase (EC 2.1.1.53).  
 GN NSPMT.  
 OS Nicotiana glauca (Wood tobacco).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 OC Asteridae; lamids; Solanales; Solanaceae; Nicotiana.  
 OX NCBI\_TaxID=4096;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=98281572; PubMed=9620262;  
 RA Hashimoto T., Shoji T., Mihara T., Oguri H., Tamaki K., Suzuki K.,  
 RT "Intraspecific variability of the tandem repeats in Nicotiana  
 RT putrescine N-methyltransferase";  
 RL Plant Mol. Biol. 37:25-37(1998).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=20419123; PubMed=10965939;  
 RA Shoji T., Yamada Y., Hashimoto T.;  
 RT "Jasmonate induction of putrescine N-methyltransferase genes in the  
 RT root of Nicotiana glauca";  
 RL Plant Cell Physiol. 41:831-839(2000).  
 DR EMBL; AB004322; BAA74542.1; -;  
 DR InterPro; IPR000051; SAM\_bind.  
 DR InterPro; IPR001045; Sprime\_synthase.  
 DR Pfam; PF01564; Spermine\_synth; 1.  
 DR TIGRPFAM; TIGR00417; spse; 1.  
 DR PROSITE; PS01330; SPERMIDINE\_SYNTHASE; 1.  
 KM Methyitransferase; Transferase.  
 SQ SEQUENCE 353 AA; 38762 MW; E083F6300F68669A CRC64;

Query Match 5.3%; Score 7; DB 10; Length 353;  
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 95 ENGFPY 101  
 Db 119 ENGFPY 125

## RESULT 37

08DSEA ID Q8DSEA4 PRELIMINARY; PRT; 354 AA.  
 AC Q8DSEA4; 01-MAR-2003 (TREMBlrel. 23, Created)  
 DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)  
 DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)  
 DE Putative aminopeptidase P.  
 GN PEPP OR SMU.1850.  
 OS Streptococcus mutans.  
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
 OC Streptococcus.  
 NCBI\_TaxID=1309;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=UAI59 / ATCC 700610 / Serotype C;  
 RX MEDLINE=22295063; PubMed=12397186;  
 RA Adic D., McShan W.M., McLaughlin R.E., Savic G., Chang J.,  
 RA Carson M.B., Primeaux C., Tian R., Kenton S., Jia H., Ian Y.,  
 RA Li S., Zhu H., Najjar F., Lai H., White J., Roe B.A., Ferrer J.J.;  
 RT "Genome sequence of Streptococcus mutans UAI59, a cariogenic dental  
 pathogen."  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:14434-14439(2002).  
 DR EMBL; AE015012; AAN59472.1; -.  
 KW Aminopeptidase; Complete proteome.  
 SQ SEQUENCE 354 AA; 39699 MW; 681730F7D30FC9C0 CRC64;

## Query Match

Best Local Similarity 100.0%; Score 7; DB 16; Length 354;  
 Pred. No. 1.1e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 106 IEAIRRA 112  
 Db 133 IEAIRRA 139

## RESULT 38

093XQ4 ID Q93XQ4 PRELIMINARY; PRT; 371 AA.  
 AC Q93XQ4; 01-DEC-2001 (TREMBlrel. 19, Created)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)  
 DT 01-OCT-2002 (TREMBlrel. 22, Last annotation update)  
 DE Putrescine N-methyltransferase 2 (EC 2.1.1.53).  
 GN PMT2.  
 OS Nicotiana attenuata.  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 OC Asteridae; lamids; Solanales; Solanaceae; Nicotiana.  
 NCBI\_TaxID=49451;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Root;  
 RX MEDLINE=21196107; PubMed=11299398;  
 RA Wenz R.A., Baldwin I.T.;  
 RT "Molecular interactions between the Specialist Herbivore Manduca sexta  
 (Lepidoptera, Spingidae) and Its Natural Host Nicotiana attenuata.  
 RT IV. Insect-Induced Ethylene Reduces Jasmonate-Induced Nicotine  
 Accumulation by Regulating Putrescine N-Methyltransferase  
 Transcripts."  
 RT Plant Physiol. 125:2189-2202(2001).  
 RL EMBL; AF280403; AAK49871.1;  
 DR InterPro; IPR000051; SAM\_bind.  
 DR InterPro; IPR001045; Sprimine synthase.  
 DR TIGR; PF01564; Spermine synth. 1.  
 DR TIGR; TIGR00417; SPEB. 1.  
 DR PROSITE; PS01330; SPERMIDINE SYNTHASE; 1.  
 KW Methyltransferase; Transaminase.  
 SQ SEQUENCE 371 AA; 40698 MW; DC6149417C19C08F CRC64;

## Query Match

5.3%; Score 7; DB 10; Length 371;

Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 95 ENGFPY 101  
 Db 137 ENGFPY 143

## RESULT 39

08K393 ID Q8K393 PRELIMINARY; PRT; 372 AA.  
 AC Q8K393; 01-OCT-2002 (TREMBlrel. 22, Created)  
 DT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)  
 DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)  
 DE RIKEN cDNA 2610307008 gene (Fragment).  
 GN 2610307008RIK.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Strausberg R.;  
 RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; BC027757; AAH27757.1; -.  
 DR MGI; MGI:1919762; 2610307008RIK.  
 FT NON TER 1  
 SQ SEQUENCE 372 AA; 42214 MW; 664528BD06E521BC CRC64;

## Query Match

Best Local Similarity 100.0%; Score 7; DB 11; Length 372;  
 Pred. No. 1.2e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 24 TSVAPP 30  
 Db 341 TSVAPP 347

## RESULT 40

09HMG7 ID Q9HMG7 PRELIMINARY; PRT; 373 AA.  
 AC Q9HMG7; 01-MAR-2001 (TREMBlrel. 16, Created)  
 DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)  
 DT 01-OCT-2002 (TREMBlrel. 22, Last annotation update)  
 DE Perichrome ABC transporter permease.  
 GN FHUG OR VNG2551G.  
 OS Halobacterium sp. (strain NRC-1).  
 OC Archaea; Euryarchaeota; Halobacteria; Halobacteriales;  
 OC Halobacteriaceae; Halobacterium.  
 NCBI\_TaxID=64091;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=20504483; PubMed=11016950;  
 RA Ng W.V., Kennedy S.P., Mahaitas G.G., Bergquist B., Pan M.,  
 RA Shukla H.D., Lasky S.R., Baliga N.S., Thorsen V., Shrogha J.,  
 RA Swartzel S., Weir D., Hall J., Dahl T.A., Welter R., Goo Y.A.,  
 RA Leitauer B., Keller K., Cruz R., Danson M.J., Hough D.W.,  
 RA Maddocks D.G., Jablonski P.E., Krebs W.P., Angevine C.M., Dale H.,  
 RA Isenbarger T.A., Beck R.F., Pohlschroder M., Spudis J.L., Jung K.-H.,  
 RA Alam M., Freitas T., Hou S., Daniels C.J., Dennis P.P., Omer A.D.,  
 RA Ehardt H., Lowe T.M., Liang P., Riley M., Hood L., DasSarma S.;  
 RT "Genome sequence of Halobacterium species NRC-1."  
 RL Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181(2000).  
 DR EMBL; AE005130; AAG20604.1; -.  
 DR InterPro; IPR000522; FecCD.  
 DR Pfam; PF01032; FecCD; 1.  
 KW Complete proteome.  
 SQ SEQUENCE 373 AA; 38785 MW; 634390E4EF649A0 CRC64;

## Query Match

5.3%; Score 7; DB 17; Length 373;  
 Best Local Similarity 100.0%; Pred. No. 1.2e+02;



Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 75 LGGTGAF 81  
 Db 163 LGGTGAF 169

RESULT 41

Q9CZY7 PRELIMINARY; PRT; 378 AA.  
 AC Q9CZY7;  
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)  
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
 DE 2610307008RIK protein.  
 GN 2610307008RIK.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Embryo;  
 RX MEDLINE=21085660; PubMed=11217851;  
 RA Kawai T., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,  
 RA Arakawa T., Hara A., Fukunishi Y., Kono H., Adachi J., Fukuda S.,  
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamakura I.,  
 RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,  
 RA Kadoya K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,  
 RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,  
 RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,  
 RA Schriml L.M., Staudl F., Suzuki R., Tomita M., Wagner L., Washio T.,  
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barab G.,  
 RA Blake J., Boffelli D., Bojunga N., Carlini P., de Bonaldo M.F.,  
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,  
 RA Gustinich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,  
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombarts P.,  
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,  
 RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,  
 RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohzuki S.,  
 RA Hayashizaki Y.,  
 RT "Functional annotation of a full-length mouse cDNA collection."  
 RL Nature 409:685-690(2001).  
 DR EMBL; AK012006; BAB27972.1; -.  
 MD: MGI:1919762; 2610307008RIK.  
 SQ SEQUENCE 378 AA; 42819 MW; 267C238CD64F5B89 CRC64;

Query Match 5.3%; Score 7; DB 11; Length 378;  
 Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 24 TSVAPP 30  
 Db 347 TSVAPP 353

RESULT 42

Q8CAZ9 PRELIMINARY; PRT; 378 AA.  
 AC Q8CAZ9;  
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE Proline arginine-rich end leucine-rich repeat.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Skin;  
 RX MEDLINE=22354683; PubMed=12466851;

RA The FANTOM Consortium,  
 RA the RIKEN Genome Exploration Research Group Phase I & II Team;  
 RT "Analysis of the mouse transcriptome based on functional annotation of  
 RT 60,770 full-length cDNAs."  
 RL Nature 420:563-573(2002).  
 DR EMBL; AK037150; BAC29722.1; -.  
 SQ SEQUENCE 378 AA; 43293 MW; 6F85A1B935FFB8C CRC64;

Query Match 5.3%; Score 7; DB 11; Length 378;  
 Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 90 VFSEKLEN 96  
 Db 186 VFSEKLEN 192

RESULT 43

Q8C227 PRELIMINARY; PRT; 378 AA.  
 AC Q8C227;  
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
 DE Hypothetical protein.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J;  
 RX MEDLINE=22354683; PubMed=12466851;  
 RA The FANTOM Consortium,  
 RA the RIKEN Genome Exploration Research Group Phase I & II Team;  
 RT "Analysis of the mouse transcriptome based on functional annotation of  
 RT 60,770 full-length cDNAs."  
 RL Nature 420:563-573(2002).  
 DR EMBL; AK089405; BAC40870.1; -.  
 KW Hypothetical protein.  
 SQ SEQUENCE 378 AA; 42829 MW; 656ED1907ACE4C8 CRC64;

Query Match 5.3%; Score 7; DB 11; Length 378;  
 Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 24 TSVAPP 30  
 Db 347 TSVAPP 353

RESULT 44

Q93X05 PRELIMINARY; PRT; 388 AA.  
 AC Q93X05;  
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)  
 DE Putrescine N-methyltransferase 1 (EC 2.1.1.53).  
 GN PMT1.  
 OS Nicotiana attenuata.  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 OC Asteridae; lamiales; Solanales; Solanaceae; Nicotiana.  
 OX NCBI\_TaxID=49451;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Root;  
 RX MEDLINE=21196107; PubMed=11299398;  
 RA Winz R.A., Baldwin I.T.,  
 RT "Molecular Interactions between the Specialist Herbivore *Manduca sexta*  
 RT (Lepidoptera, Sphingidae) and its Natural Host *Nicotiana attenuata*.  
 RT IV. Insect-Induced Ethylene Reduces Jasmonate-Induced Nicotine

```

RT Accumulation by Regulating Putrescine N-Methyltransferase
RT Transcripts: 1.
RL Plant Physiol. 125:2189-2202(2001).
DR EMBL: AP280402; AAK49870.1; -.
DR InterPro: IPR000051; SAM_Bind.
DR InterPro: IPR001045; Sptamine_synthase.
DR Pfam: PF01564; Spermine synth. 1.
DR TRAFAM: TIGR00417; spee. 1.
DR PROSITE: PS01330; SPERMIDINE SYNTHASE; 1.
KM Methyltransferase; Transferrase.
SQ SEQUENCE 388 AA; 42627 MW; 413B8C1745BA2925 CRC64;

Query Match 5.3%; Score 7; DB 10; Length 388;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 95 ENGCFPY 101
DB 154 ENGCFPY 160

RESULT 45
Q9RJD0 PRELIMINARY; PRT; 389 AA.
ID Q9RJD0;
AC Q9RJD0;
DT 01-MAY-2000 (TReMBLrel. 13, Created)
DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)
DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)
DE Putative oxidoreductase.
GN SC00763 OR SCF81.22.
OS Streptomyces coelicolor.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycineae; Streptomycetaceae; Streptomycetes.
NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Oliver K., Harris D.;
RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Cardeno A.M., Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RX MEDLINE=97000351; PubMed=8843436;
RA Redenbach M., Kleiser H.M., Denapate D., Eichner A., Cullum J.,
RA Knaeht H., Hopwood D.A.;
RT "A set of ordered cosmids and a detailed genetic and physical map for
RT the 8 Mb Streptomyces coelicolor A3(2) chromosome."
RL Mol. Microbiol. 21:77-96(1996).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2) / M145;
RX MEDLINE=9196410; PubMed=12000953;
RA Bentley S.D., Chater K.F., Cerdano-Iarraga A.-M., Challis G.L.,
RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kleiser H.,
RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA Huang C.-H., Kleiser T., Larke L., Murphy L., Oliver K., O'Neill S.,
RA Rabbittsch E., Rajandream M.A., Rutherford K., Ruter S.,
RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
RA Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,
RA Hopwood D.A.;
RT "Complete genome sequence of the model actinomycete Streptomyces
RT coelicolor A3(2).";
RL Nature 417:141-147(2002).
DR EMBL: AL393106; CAB61541.1; -.
DR HSSP; P00175; ILTD.
DR InterPro: IPR003009; FMN_enzyme.
DR InterPro: IPR000262; FMN_hydxyc_dh.

```

```

DR Pfam: PF01070; FMN_dh. 1.
DR PROSITE: PS00557; FMN_HYDROXY_ACID_DH; 1.
KM Complete proteome.
SQ SEQUENCE 389 AA; 41742 MW; DCDABACC6E2F56FA CRC64;

Query Match 5.3%; Score 7; DB 16; Length 389;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 PEAEPAA 16
DB 108 PEAEPAA 114

RESULT 46
Q9VIM3 PRELIMINARY; PRT; 418 AA.
ID Q9VIM3;
AC Q9VIM3;
DT 01-MAY-2000 (TReMBLrel. 13, Created)
DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)
DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)
DE CG9797 protein (LD30467P).
GN CG9797.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brannon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Aghayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borokova D., Botchan P.V., Bouck J., Brokstein P., Brotter P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasner K.,
RA Glodet A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Kapen G.H., Ke Z., Kension J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Paclebo J.M.,
RA Palazolo M., Peltan G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Sidgen-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Waseraman D.A., Weinstock G.M., Weisenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers B.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkeley;

```

RA Stapleton M., Brokstein P., Hong L., Adbayani A., Carlson J.,  
 RA Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,  
 RA Gonzalez M., Garin H., Li P., Liao G., Miranda A., Mungall C.J.,  
 RA Nuno J., Pacleb J., Paragas V., Park S., Phouanavong S., Wan K.,  
 RA Yu C., Lewis S.E., Rubin G.M., Celis S.,  
 RL Submitted (Aug-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AB003680; AAF54280.1; -  
 DR EMBL; AY051784; AAF53208.1; -  
 DR FlyBase; FBgn0037621; CG9797.  
 DR InterPro; IPR007087; Znf C2H2.  
 DR Pfam; PF00096; zfc2h2; 5.  
 DR SMART; SM00355; Znf C2H2; 5.  
 DR PROSITE; PS00028; ZINC\_FINGER\_C2H2\_1; 5.  
 DR PROSITE; PS0157; ZINC\_FINGER\_C2H2\_2; 5.  
 KW Metal-binding; Zinc; Zinc-finger.  
 SQ SEQUENCE 418 AA; 48093 MW; 28EA00F1F455FAD0 CRC64;

Query Match 5.3%; Score 7; DB 5; Length 418;  
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 57 ELASAVK 63  
 Db 65 ELASAVK 71

RESULT 47  
 O97W60 PRELIMINARY; PRT; 421 AA.  
 AC O97W60;  
 DT 01-OCT-2001 (TREMBlrel. 18, Created)  
 DT 01-OCT-2001 (TREMBlrel. 18, Last sequence update)  
 DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)  
 DE Hypochemical protein SSO2380.  
 GN SSO2380.  
 OS Sulfolobus solfataricus.  
 OC Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;  
 OC Sulfolobus.  
 OX NCBI\_TaxId=2287;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 35092 / DSM 1617 / P2;  
 RX MEDLINE=21332296; PubMed=11427726;  
 RA She Q., Singh R.K., Confalonieri F., Zivanovic Y., Allard G.,  
 RA Awayshe M.J., Chan-Weher C.C.-Y., Clausen I.G., Curtis B.A.,  
 RA De Moors A., Krause G., Fletcher C., Gordon P.M.K.,  
 RA Heikamp-de Jong I., Jeffries A.C., Kozera C.J., Medina N., Peng X.,  
 RA Thi Ngoc H.P., Redder P., Schenk M.E., Theriault C., Tolstrup N.,  
 RA Charlebois R.L., Doolittle W.F., Duguet M., Gaasterland T.,  
 RA Garrett R.A., Ragan M.A., Sensen C.W., Van der Oost J.,  
 RA "The complete genome of the crenarchaeon Sulfolobus solfataricus P2.",  
 RL Proc. Natl. Acad. Sci. U.S.A. 98:7835-7840(2001).  
 DR EMBL; AB006839; AAK42528.1; -  
 DR InterPro; IPR006638; ELP3.  
 DR InterPro; IPR006466; M1ab\_1like\_B.  
 DR InterPro; IPR002792; TRAM.  
 DR InterPro; IPR005839; UPF0004.  
 DR Pfam; PF01938; TRAM; 1.  
 DR Pfam; PR00919; UPF0004; 1.  
 DR SMART; SM00729; ELP3; 1.  
 DR TIGRFAms; TIGR01578; M1ab\_1like-B; 1.  
 DR TIGRFAms; TIGR00089; TIGR00089; 1.  
 KW Hypochemical protein; Complete proteome.  
 SQ SEQUENCE 421 AA; 47262 MW; 9F6D78AD435927C9 CRC64;

Query Match 5.3%; Score 7; DB 17; Length 421;  
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 104 DLIEAIR 110  
 Db 232 DLIEAIR 238

RESULT 48  
 O9ZWT7 PRELIMINARY; PRT; 430 AA.  
 AC O9ZWT7;  
 DT 01-MAY-1999 (TREMBlrel. 10, Created)  
 DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)  
 DT 01-OCT-2002 (TREMBlrel. 22, Last annotation update)  
 DE Putrescine N-methyltransferase.  
 GN NSPMT3.  
 OS Nicotiana glauca (Wood tobacco).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 OC Asteridae; lamiales; Solanales; Solanaceae; Nicotiana.  
 OX NCBI\_TaxId=4096;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=98281572; PubMed=9620262;  
 RA Hashimoto T., Shoji T., Mihara T., Oguri H., Tamaki K., Suzuki K.,  
 RA Yamada Y.,  
 RL "Intraspecific variability of the tandem repeats in Nicotiana  
 RL putrescine N-methyltransferase.",  
 RL Plant Mol. Biol. 37:25-37(1998).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=20419123; PubMed=10965939;  
 RA Shoji T., Yamada Y., Hashimoto T.,  
 RT "Jaomonte induction of Putrescine N-methyltransferase genes in the  
 RT root of Nicotiana glauca.",  
 RL Plant Cell Physiol. 41:831-839(2000).  
 DR EMBL; AB004324; BAA74544.1; -  
 DR InterPro; IPR000051; SAM\_bind.  
 DR InterPro; IPR001045; Spermine synthase.  
 DR Pfam; PF01564; Spermine synth; 1.  
 DR TIGRFAms; TIGR00417; spee; 1.  
 DR PROSITE; PS01330; SPERMIDINE\_SYNTHASE; 1.  
 KW Methyltransferase; Transferase.  
 SQ SEQUENCE 430 AA; 47178 MW; ADEA1BEC73A4CC2 CRC64;

Query Match 5.3%; Score 7; DB 10; Length 430;  
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 95 ENGPFY 101  
 Db 196 ENGPFY 202

RESULT 49  
 O9KZ58 PRELIMINARY; PRT; 438 AA.  
 AC O9KZ58;  
 DT 01-OCT-2000 (TREMBlrel. 15, Created)  
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
 DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)  
 DE Putative fatty acid synthase.  
 GN SCO3091 OR SCE25.32C.  
 OS Streptomyces coelicolor.  
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
 OC Streptomycinae; Streptomycetaceae; Streptomyces.  
 OX NCBI\_TaxId=1902;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=A3(2);  
 RA Seeger K.J., Harris D.,  
 RL Submitted (Apr-2000) to the EMBL/GenBank/DBJ databases.  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=A3(2);  
 RA Cerdano A.M., Parkhill J., Barrell B.G., Rajandream M.A.,  
 RL Submitted (Apr-2000) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.

RC STRAIN=A3(2);  
 RX MEDLINE=97000351; PubMed=8843436;  
 RA Redenbach M., Kleser H.M., Denapaite D., Eichner A., Cullum J.,  
 RA Kinashi H., Hopwood D.A.,  
 RT "A set of ordered cosmids and a detailed genetic and physical map for  
 the 8 Mb Streptomyces coelicolor A3(2) chromosome.";  
 RL Mol. Microbiol. 21:77-96(1996).  
 RN [4]

RP SEQUENCE FROM N.A.  
 RC STRAIN=A3(2) / M145;  
 RX MEDLINE=21996410; PubMed=12000953;  
 RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,  
 RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kleser H.,  
 RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,  
 RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornaby T., Howarth S.,  
 RA Huang C.H., Kleser T., Larke L., Murphy L., Oliver K., O'Neil S.,  
 RA Rabbittowitch E., Rajandream M.A., Rutherford K., Rutter S.,  
 RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,  
 RA Warren T., Wierozorek A., Woodward J., Barrell B.G., Parkhill J.,  
 RA Hopwood D.A.;  
 RT "Complete genome sequence of the model actinomycete Streptomyces  
 coelicolor A3(2)."  
 RL Nature 417:141-147(2002).  
 DR EMBL; AL939115; CAB89463.1; -.  
 DR InterPro; IPR003333; CMAS.  
 DR InterPro; IPR001601; Methyitransf.  
 DR InterPro; IPR000051; SAM\_bind.  
 DR Pfam; PF02353; CMAS; 1.  
 KW Complete proteome.  
 SQ SEQUENCE 438 AA; 48190 MW; CB8BF7C9169B3ADB CRC64;

Query Match 5.3%; Score 7; DB 16; Length 438;  
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 27 APPPEEV 33  
 |||||  
 Db 123 APPPEEV 129

RESULT 50  
 Q8YX47 PRELIMINARY; PRT; 454 AA.  
 AC Q8YX47;  
 DT 01-MAR-2002 (TrEMBLrel. 20, Created)  
 DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)  
 DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)  
 DE Hypothetical protein ALR1370.  
 GN ALR1370.  
 OS Anabaena sp. (strain PCC 7120).  
 OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.  
 OX NCBI\_TaxID=103690;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=21595285; PubMed=11759840;  
 RA Kaneko T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,  
 RA Matsumoto M., Ishikawa A., Kawashima K., Kimura T.,  
 RA Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,  
 RA Nakazaki N., Shimpo S., Sugimoto M., Takazawa M., Yamada M.,  
 RA Yasuda M., Tabata S.;  
 RT "Complete genomic sequence of the filamentous nitrogen-fixing  
 cyanobacterium Anabaena sp. strain PCC 7120.";  
 RL DNA Res. 8:205-213(2001).  
 DR EMBL; AF003585; BAB73327.1; -.  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 454 AA; 49858 MW; 5B6A5592CC793A81 CRC64;

Db 17 APPPEEV 23  
 Search completed: December 15, 2003, 17:06:50  
 Job time : 40 secs

Query March 5.3%; Score 7; DB 16; Length 454;  
 Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 27 APPPEEV 33  
 |||||

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

## OM protein - protein search, using sw model

Run on: December 15, 2003, 16:59:56 ; Search time 41 Seconds  
(without alignments)  
507.151 Million cell updates/sec

Title: US-09-925-301-966

Perfect score: 131

Sequence: 1 AEVHTRKQGEAPAAAMSGE.....ASNGETLEKITSRPPCVIL 131

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 1107863 seqs, 158726573 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 100 summaries

Database :

A\_Geneseq\_19Jun03:\*

1: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:\*  
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:\*  
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:\*  
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:\*  
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:\*  
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:\*  
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:\*  
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:\*  
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:\*  
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:\*  
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:\*  
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:\*  
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:\*  
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:\*  
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:\*  
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:\*  
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:\*  
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:\*  
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:\*  
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:\*  
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:\*  
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:\*  
23: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:\*  
24: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2003.DAT:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	131	100.0	131	21	AA843521 Human cancer assoc
2	120	91.6	124	23	ABP43055 Human ovarian anti
3	120	91.6	206	24	ABR47619 Breast cancer asso
4	115	87.8	115	22	AA878970 Human C35 protein.
5	115	87.8	115	22	AA878997 Human C35, a tumor
6	115	87.8	115	23	ABP58560 Human site-specifi
7	115	87.8	115	23	ABP43843 RIKEN 1810045019 p
8	27	20.6	90	21	AA803153 Human secreted pro
9	21	16.0	21	22	AA878988 Human C35 peptide

10	21	16.0	21	22	AA87899	Human C35 peptide
11	21	16.0	21	22	ABBI4673	Human C35 peptide
12	21	16.0	21	22	ABBI4674	Human C35 peptide
13	19	14.5	19	22	AA877901	Human C35 peptide
14	19	14.5	19	22	ABBI4672	Human C35 peptide
15	15	11.5	15	22	AA877900	Human C35 peptide
16	15	11.5	15	22	ABBI2825	Human C35 peptide
17	15	11.5	15	22	ABBI2826	Human C35 peptide
18	15	11.5	15	22	ABBI2827	Human C35 peptide
19	15	11.5	15	22	ABBI2828	Human C35 peptide
20	15	11.5	15	22	ABBI2829	Human C35 peptide
21	15	11.5	15	22	ABBI2830	Human C35 peptide
22	15	11.5	15	22	ABBI2831	Human C35 peptide
23	15	11.5	15	22	ABBI2832	Human C35 peptide
24	15	11.5	15	22	ABBI2833	Human C35 peptide
25	15	11.5	15	22	ABBI2834	Human C35 peptide
26	15	11.5	15	22	ABBI2835	Human C35 peptide
27	15	11.5	15	22	ABBI2836	Human C35 peptide
28	15	11.5	15	22	ABBI2837	Human C35 peptide
29	15	11.5	15	22	ABBI2838	Human C35 peptide
30	15	11.5	15	22	ABBI2839	Human C35 peptide
31	15	11.5	15	22	ABBI2840	Human C35 peptide
32	15	11.5	15	22	ABBI2841	Human C35 peptide
33	15	11.5	15	22	ABBI2842	Human C35 peptide
34	15	11.5	15	22	ABBI2843	Human C35 peptide
35	15	11.5	15	22	ABBI2844	Human C35 peptide
36	15	11.5	15	22	ABBI2845	Human C35 peptide
37	15	11.5	15	22	ABBI2846	Human C35 peptide
38	15	11.5	15	22	ABBI2847	Human C35 peptide
39	15	11.5	15	22	ABBI2848	Human C35 peptide
40	15	11.5	15	22	ABBI2849	Human C35 peptide
41	15	11.5	15	22	ABBI2850	Human C35 peptide
42	15	11.5	15	22	ABBI2851	Human C35 peptide
43	15	11.5	15	22	ABBI2852	Human C35 peptide
44	15	11.5	15	22	ABBI2853	Human C35 peptide
45	15	11.5	15	22	ABBI2854	Human C35 peptide
46	15	11.5	15	22	ABBI2855	Human C35 peptide
47	15	11.5	15	22	ABBI2856	Human C35 peptide
48	15	11.5	15	22	ABBI2857	Human C35 peptide
49	15	11.5	15	22	ABBI2858	Human C35 peptide
50	15	11.5	15	22	ABBI2859	Human C35 peptide
51	15	11.5	15	22	ABBI2860	Human C35 peptide
52	15	11.5	15	22	ABBI2861	Human C35 peptide
53	15	11.5	15	22	ABBI2862	Human C35 peptide
54	15	11.5	15	22	ABBI2863	Human C35 peptide
55	15	11.5	15	22	ABBI2864	Human C35 peptide
56	15	11.5	15	22	ABBI2865	Human C35 peptide
57	15	11.5	15	22	ABBI2866	Human C35 peptide
58	15	11.5	15	22	ABBI2867	Human C35 peptide
59	15	11.5	15	22	ABBI2868	Human C35 peptide
60	15	11.5	15	22	ABBI2869	Human C35 peptide
61	15	11.5	15	22	ABBI2870	Human C35 peptide
62	15	11.5	15	22	ABBI2871	Human C35 peptide
63	15	11.5	15	22	ABBI2872	Human C35 peptide
64	15	11.5	15	22	ABBI2873	Human C35 peptide
65	15	11.5	15	22	ABBI2874	Human C35 peptide
66	15	11.5	15	22	ABBI2875	Human C35 peptide
67	15	11.5	15	22	ABBI2876	Human C35 peptide
68	15	11.5	15	22	ABBI2877	Human C35 peptide
69	15	11.5	15	22	ABBI2878	Human C35 peptide
70	15	11.5	15	22	ABBI2879	Human C35 peptide
71	15	11.5	15	22	ABBI2880	Human C35 peptide
72	15	11.5	15	22	ABBI2881	Human C35 peptide
73	15	11.5	15	22	ABBI2882	Human C35 peptide
74	15	11.5	15	22	ABBI2883	Human C35 peptide
75	15	11.5	15	22	ABBI2884	Human C35 peptide
76	15	11.5	15	22	ABBI2885	Human C35 peptide
77	15	11.5	15	22	ABBI2886	Human C35 peptide
78	15	11.5	15	22	ABBI2887	Human C35 peptide
79	15	11.5	15	22	ABBI2888	Human C35 peptide
80	15	11.5	15	22	ABBI2889	Human C35 peptide
81	15	11.5	15	22	ABBI2890	Human C35 peptide
82	15	11.5	15	22	ABBI2891	Human C35 peptide

83	15	11.5	15	22	ABB12892	Human C35 peptide
84	15	11.5	15	22	ABB12893	Human C35 peptide
85	15	11.5	15	22	ABB12894	Human C35 peptide
86	15	11.5	15	22	ABB12895	Human C35 peptide
87	15	11.5	15	22	ABB12896	Human C35 peptide
88	15	11.5	15	22	ABB12897	Human C35 peptide
89	15	11.5	15	22	ABB12898	Human C35 peptide
90	15	11.5	15	22	ABB12899	Human C35 peptide
91	15	11.5	15	22	ABB12900	Human C35 peptide
92	15	11.5	15	22	ABB12901	Human C35 peptide
93	15	11.5	15	22	ABB12902	Human C35 peptide
94	15	11.5	15	22	ABB12903	Human C35 peptide
95	15	11.5	15	22	ABB12904	Human C35 peptide
96	15	11.5	15	22	ABB12905	Human C35 peptide
97	15	11.5	15	22	ABB12906	Human C35 peptide
98	15	11.5	15	22	ABB12907	Human C35 peptide
99	15	11.5	15	22	ABB12908	Human C35 peptide
100	15	11.5	15	22	ABB12909	Human C35 peptide

## ALIGNMENTS

RESULT 1  
AAB43521  
ID AAB43521 standard; Protein; 131 AA.

AC AAB43521;

DT 08-FEB-2001 (first entry)

DE Human cancer associated protein sequence SEQ ID NO:966.

XX Human cancer associated gene; cancer antigen; detection; cancer;  
KM diagnosis; cytostatic; proliferative; vulnerrary; immunomodulator;  
KM antidiabetic; antisthmatic; antirheumatic; antiarthritic; antiviral;  
KM antidiabetic; antithyroid; antiallergic; antibacterial; cardiant;  
KM dermatological; neuroprotective; thrombolytic; coagulant; neutropic;  
KM vasotropic; antipruritic; antidiogenic; gene therapy; inflammation;  
KM immune disorder; haematopoietic cell disorder; autoimmune disorder;  
KM allergic reaction; graft versus host disease; organ rejection;  
KM haemostatic; thrombolytic; cardiovascular disorder; infection;  
KM neurological disease; drug screening.

OS Homo sapiens.

PN WO200055350-A1.

XX 21-SEP-2000.

PF 08-MAR-2000; 2000WO-US05882.

PR 12-MAR-1999; 99US-0124270.

PA (HUMA-) HUMAN GENOME SCI INC.

PI Rosen CA, Ruben SM;

XX WPI, 2000-587533/55.

DR N-PSDB; AAC77730.

XX Novel isolated nucleic acids comprising sequences encoding peptides

PT useful for treating or diagnosing e.g. cancer -

XX Claim 11; Page 1534-1535; 2352pp; English.

XX AAC77607 to AAC78448 encode the human cancer associated proteins given  
CC in AAB43398 to AAB44239. The proteins can have activities based on the  
CC tissues and cells the genes are expressed in. Example of activities  
CC include: cytostatic; proliferative; vulnerrary; immunomodulator;  
CC antidiabetic; antisthmatic; antirheumatic; antiarthritic;  
CC antidiabetic; antithyroid; antiallergic; antibacterial; antiviral;  
CC dermatological; neuroprotective; cardiant; thrombolytic; coagulant;

CC neutropic; vasotropic; antipruritic and antidiogenic. The  
CC polynucleotides and polypeptides can be used for preventing, treating or  
CC ameliorating medical conditions and diagnosing pathological conditions.  
CC Polynucleotides, polypeptides, antibodies, agonists and antagonists from  
CC the present invention may be used to treat immune disorders by activating  
CC or inhibiting the proliferation, differentiation or mobilisation of  
CC immune cells, to treat disorders of haematopoietic cells, autoimmune  
CC disorders, allergic reactions, graft versus host disease and organ  
CC rejection, modulate haemostatic or thrombolytic activity, modulate  
CC inflammation, cancers, cardiovascular disorders, neurological disease and  
CC bacterial or viral infections. The peptides, nucleotides, antibodies,  
CC agonists and antagonists may be also used in drug screens. AAC78449 to  
CC AAC78457 and AAB44240 represent sequences used in the exemplification of  
CC the present invention.

XX Sequence 131 AA;

Query Match 100.0%; Score 131; DB 21; Length 131;

Best Local Similarity 100.0%; Pred. No. 1.7e-120; Indels 0; Gaps 0;

Matches 131; Conservative 0; Mismatches 0;

QY	1	AEVHTRKQGEAPAMSGEPQTSVAPPEEVEPGSGVRIWYCEPGCFATYELAS	60
DB	1	AEVHTRKQGEAPAMSGEPQTSVAPPEEVEPGSGVRIWYCEPGCFATYELAS	60
QY	61	AVKEQYFGIEIESRLGTAFFIEIINGQLVFSKLENGCFPYEKDLEAIRRANGSTLEK	120
DB	61	AVKEQYFGIEIESRLGTAFFIEIINGQLVFSKLENGCFPYEKDLEAIRRANGSTLEK	120
QY	121	ITNSRPPCVIL	131
DB	121	ITNSRPPCVIL	131

RESULT 2

ABP43055  
ID ABB43055 standard; Protein; 124 AA.

XX ABP43055;

DT 22-AUG-2002 (first entry)

DE Human ovarian antigen HVC468, SEQ ID NO:4187.

XX Human; ovarian antigen; ovary; ovarian; breast; cancer; tumour;  
KM ovarian cancer; breast cancer; tumour; reproductive system disorder;  
KM infertility; pregnancy disorder; anovulation; polycystic ovary syndrome;  
KM PCOS; ovarian cyst; dysmenorrhoea; endocrine disorder; infection;  
KM inflammatory condition; immune disorder; blood disorder;  
KM cardiovascular disorder; respiratory disorder; neurological disorder;  
KM gastrointestinal disorder; urinary system disorder; drug screening;  
KM gene therapy; chromosome mapping; forensic analysis;  
KM antibody preparation; cytostatic; immunomodulatory; neuroprotective;  
KM antidiabetic; gynaecological; reproductive.

OS Homo sapiens.

PN WO200200677-A1.

XX 03-JAN-2002.

PF 07-JUN-2001; 2001WO-US18569.

PR 07-JUN-2000; 2000US-209467P.

PA (HUMA-) HUMAN GENOME SCI INC.

PI Birse CE, Rosen CA;

XX WPI, 2002-147878/19.

DR N-PSDB; ABQ56133.

PT Isolated nucleic acid molecules encoding novel ovarian polypeptides,

PT useful in the prevention, treatment and diagnosis of cancer (e.g.  
PT ovarian cancer), immune disorders, cardiovascular disorders and  
PT neurological diseases -

PS Claim 11, SEQ ID No 4187; 2922pp; English.

XX The invention relates to 2175 novel human ovarian antigens (ABP41054-  
CC ABP43228) and to cDNAs encoding them (ABQ54131-ABQ56305), and also  
CC encompasses polypeptides 90% identical and polynucleotides 95% identical  
CC to the sequences of the invention. The invention additionally relates to  
CC recombinant vectors and host cells comprising human ovarian antigen  
CC polynucleotides, antibodies against human ovarian antigens, and the use  
CC of ovarian antigen polynucleotides and polypeptides in diagnosing,  
CC treating, prognosing or preventing various ovarian and/or breast-related  
CC disorders. Such conditions include ovarian cancer and breast cancer, and  
CC metastatic tumours of ovarian or breast origin, reproductive system  
CC disorders (e.g., infertility, disorders of pregnancy, anovulation,  
CC polycystic ovary syndrome, ovarian cysts, and dysmenorrhoea), endocrine  
CC disorders, infections (e.g., chlamydia, HIV, toxoplasmosis, and toxic  
CC shock syndrome), inflammatory conditions (e.g., mastitis, oophoritis and  
CC vaginitis), immune disorders (e.g., congenital and acquired  
CC immunodeficiencies, autoimmune oophoritis, systemic lupus erythematosus),  
CC blood-related disorders (e.g., anaemia), cardiovascular disorders,  
CC respiratory disorders, neurological disorders, gastrointestinal disorders  
CC and urinary system disorders. Ovarian antigen polypeptides and  
CC polynucleotides may also be used in screening for compounds which  
CC modulate ovarian antigen expression or activity. The polynucleotides may  
CC further be used for gene therapy, chromosome mapping, in the  
CC identification of individuals and in forensic analysis, and the  
CC polypeptides may be used as food additives or to prepare antibodies  
CC useful in disease diagnosis, drug targeting and phenotyping. The present  
CC sequence represents a human ovarian antigen of the invention.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 124 AA;

Query Match 91.6%; Score 120; DB 23; Length 124;  
Best Local Similarity 100.0%; Pred. No. 1e-109;  
Matches 120; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 AEPAMSGPGQTSVAPPEEVEPGSGVRIVVEYCEPCGFATYELASAVKEQYGIET 71  
DB 5 AEPAMSGPGQTSVAPPEEVEPGSGVRIVVEYCEPCGFATYELASAVKEQYGIET 64  
QY 72 ESRIGGTGAFETIENGQVFSKLENGSPYKDLIARIRASNGTLEKITSRPPCVIL 131  
DB 65 ESRIGGTGAFETIENGQVFSKLENGSPYKDLIARIRASNGTLEKITSRPPCVIL 124

RESULT 3

ID ABR47619 standard; Protein; 206 AA.

AC ABR47619;

DT 12-JUN-2003 (first entry)

DE Breast cancer associated protein sequence SEQ ID NO:480.

XX Human; breast cancer; cytosolic; gene therapy.

OS Homo sapiens.

XX WO2003004989-A2.

XX 16-JAN-2003.

XX 21-JUN-2002; 2002WO-US19669.

XX 21-JUN-2001; 2001US-299887P.

XX 27-JUN-2001; 2001US-301572P.

PR 18-JUL-2001; 2001US-306501P.  
PR 25-SEP-2001; 2001US-325002P.  
PR 05-MAR-2002; 2002US-362585P.  
PR 14-MAY-2002; 2002US-380391P.

XX (MILL-) MILLENIUM PHARM INC.

XX Lilie J, Gannavarapu M, Glatt K, Hoersch S, Kamatkar S, Mertens M,  
PI Monahan JE, Myer V, Wang Y, Xu Y, Zhao X, Meyers RE, Bast RC;  
PI Horobagayi GN, Pusztai L, Meric F, Sahin A, Mills GB;

XX WPI, 2003-210381/20.  
DR N-PSDB; ACC50321.

PT Breast cancer diagnosis or treatment by comparing the level of  
PT expression of a marker in a patient sample with that in the control  
PT non-breast cancer sample -

PS Claim 1; SEQ ID 480; 128pp; English.

XX The present invention describes a method for assessing whether a patient  
CC is afflicted with breast cancer. The method comprises comparing the level  
CC of expression of a marker (gene/polypeptide see ACC50076 to ACC50334 and  
CC ABR47386 to ABR47632) in a patient sample and the normal level of  
CC expression of the marker in a control non-breast cancer sample, where a  
CC significant increase in the level of expression of the marker in the  
CC patient sample and the normal level is an indication that the patient is  
CC afflicted with breast cancer. The breast cancer associated sequences  
CC from the present invention have cytosolic activities and can be used in  
CC gene therapy. The method is useful for diagnosing and treating breast  
CC cancer.  
CC N.B. The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 206 AA;

Query Match 91.6%; Score 120; DB 24; Length 206;  
Best Local Similarity 100.0%; Pred. No. 1.6e-109;  
Matches 120; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 AEPAMSGPGQTSVAPPEEVEPGSGVRIVVEYCEPCGFATYELASAVKEQYGIET 71  
DB 87 AEPAMSGPGQTSVAPPEEVEPGSGVRIVVEYCEPCGFATYELASAVKEQYGIET 146  
QY 72 ESRIGGTGAFETIENGQVFSKLENGSPYKDLIARIRASNGTLEKITSRPPCVIL 131  
DB 147 ESRIGGTGAFETIENGQVFSKLENGSPYKDLIARIRASNGTLEKITSRPPCVIL 206

RESULT 4

ID AAG77870 standard; Protein; 115 AA.

AC AAG77870;

DT 08-MAY-2002 (first entry)

DE Human C35 protein.

XX C35 protein; antigenic peptide; major histocompatibility complex;  
KW MHC-peptide complex; MHC; human; MHC class I alpha chain;  
KW beta-2 microglobulin; MHC class II alpha chain; MHC class II beta chain;  
KW vaccine; immune response modulation; hyperproliferative disorder;  
KW neoplasia; hypergammaglobulinaemia; viral infection; hepatitis;  
KW meningitis; bacterial infection; tuberculosis; gingivitis;  
KW parasitic infection; autoimmune disease; Hashimoto's disease;  
KW Graves' disease; rheumatoid arthritis; allergy; asthma; organ rejection;  
KW graft-versus-host disease; GVHD; breast cancer.

XX Homo sapiens.

XX WO200178768-A2.

XX 25-OCT-2001.  
 PD 12-APR-2001; 2001WO-US1912.  
 XX 12-APR-2000; 2000US-196472P.  
 PF 12-APR-2000; 2000US-196472P.  
 XX (UVRP ) UNIV ROCHESTER.  
 XX Zauderer M, Smith ES;  
 XX WPI; 2001-602927/68.  
 DR N-PSDB; AAH77148.  
 XX Novel compound comprising major histocompatibility complex-peptide  
 PT complexes, used to modulate immune responses -  
 PS Example 16; Fig 7; 166pp; English.  
 XX The invention comprises a compound which contains one or more major  
 CC histocompatibility complex (MHC)-peptide complexes, and an antibody  
 CC specific for a cell surface marker. The complexes comprise an MHC class  
 CC I alpha chain, a beta-2 microglobulin molecule and an antigenic peptide  
 CC bound in the MHC groove. Alternatively, the complexes may comprise an MHC  
 CC class II alpha chain, an MHC class II beta chain, and an antigenic  
 CC peptide bound in the MHC groove. The complexes are linked to the carboxyl  
 CC terminus of the antibody. The compounds of the invention can be used as a  
 CC vaccine to modulate an immune response. The compounds of the invention  
 CC are useful for treating: hyperproliferative disorders (e.g. neoplasms and  
 CC hypergammaglobulinemia); viral infections (e.g. hepatitis and  
 CC meningitis); bacterial infections (e.g. tuberculosis and gingivitis);  
 CC parasitic infections; autoimmune diseases (e.g. Hashimoto's disease,  
 CC Graves' disease and rheumatoid arthritis); allergic reactions/conditions  
 CC (e.g. asthma). The compounds of the invention may also be used in the  
 CC treatment of organ rejection or graft-versus-host disease (GvHD). The  
 CC present sequence represents C35 protein, this protein is differentially  
 CC expressed in human breast cancer cells. In one embodiment of the  
 CC invention, antigenic peptides derived from the C35 protein can be used to  
 CC promote an immune response against a cancerous cell.  
 CC  
 SQ Sequence 115 AA;  
 Query Match 87.8%; Score 115; DB 22; Length 115;  
 Best Local Similarity 100.0%; Pred. No. 7.7e-105; Indels 0; Gaps 0;  
 Matches 115; Conservative 0; Mismatches 0;  
 QY 17 MSGEPQGTVAAPPEEVEPGSGVRIVVEYCEPGFEATYLELASAVKEQYPGIEISRLG 76  
 DB 1 MSGEPQGTVAAPPEEVEPGSGVRIVVEYCEPGFEATYLELASAVKEQYPGIEISRLG 60  
 QY 77 GTGAFEIENGQVFSKLENGGFPYKDLIEAIRRANGTLEKITSRPPCVIL 131  
 DB 61 GTGAFEIENGQVFSKLENGGFPYKDLIEAIRRANGTLEKITSRPPCVIL 115  
 RESULT 5  
 AAG78997  
 ID AAG78997 standard; Protein; 115 AA.  
 XX AAG78997;  
 AC 22-JAN-2002 (first entry)  
 XX Human C35, a tumour antigen.  
 DE Human C35, a tumour antigen.  
 XX Human C35; cytostatic; gene therapy; vaccine; tumour antigen;  
 KM breast cancer; bladder cancer; tumour immunotherapy; chromosome 17q12.  
 XX Homo sapiens.  
 OS Homo sapiens.  
 XX WO200174859-A2.  
 PN 11-OCT-2001.

XX 04-APR-2001; 2001WO-US10855.  
 PF 04-APR-2000; 2000US-194463P.  
 PR 04-APR-2000; 2000US-194463P.  
 XX (UVRP ) UNIV ROCHESTER.  
 XX Zauderer M, Evans EE, Borrello MA;  
 XX WPI; 2001-626383/72.  
 DR N-PSDB; AAI71785, AAI71793.  
 XX Novel C35 polypeptides and C35 genes useful in immunogenic compositions  
 PT and vaccines, for inducing antibody and cell-mediated immunity against  
 PT target cells, such as tumor cells that express C35 gene -  
 XX Claim 11; Fig 1; 33pp; English.  
 XX The present sequence is human C35. C35 is a novel tumour antigen that is  
 CC overexpressed in human breast and bladder carcinoma. C35 is thought to be  
 CC a promising candidate for tumour immunotherapy, in immunogenic  
 CC compositions and vaccines, to induce antibody and cell-mediated immunity  
 CC against target cells such as tumour cells that express C35 genes. The C35  
 CC gene aligns on human chromosome 17q12.  
 CC  
 SQ Sequence 115 AA;  
 Query Match 87.8%; Score 115; DB 22; Length 115;  
 Best Local Similarity 100.0%; Pred. No. 7.7e-105; Indels 0; Gaps 0;  
 Matches 115; Conservative 0; Mismatches 0;  
 QY 17 MSGEPQGTVAAPPEEVEPGSGVRIVVEYCEPGFEATYLELASAVKEQYPGIEISRLG 76  
 DB 1 MSGEPQGTVAAPPEEVEPGSGVRIVVEYCEPGFEATYLELASAVKEQYPGIEISRLG 60  
 QY 77 GTGAFEIENGQVFSKLENGGFPYKDLIEAIRRANGTLEKITSRPPCVIL 131  
 DB 61 GTGAFEIENGQVFSKLENGGFPYKDLIEAIRRANGTLEKITSRPPCVIL 115  
 RESULT 6  
 ABP58560  
 ID ABP58560 standard; Protein; 115 AA.  
 XX ABP58560;  
 AC 28-MAR-2003 (first entry)  
 DT Human site-specific recombinase motif-containing protein 12.63.  
 DE Human site-specific recombinase motif-containing protein 12.63.  
 XX Human, site-specific recombinase motif-containing protein 12.63;  
 KM recombinant production; gene therapy; cancer; tumour; HIV infection;  
 KW human immunodeficiency virus; cytostatic.  
 XX Homo sapiens.  
 OS Homo sapiens.  
 XX CN1363589-A.  
 PN 14-AUG-2002.  
 PD 05-JAN-2001; 2001CN-0105072.  
 PF 05-JAN-2001; 2001CN-0105072.  
 XX 05-JAN-2001; 2001CN-0105072.  
 PR (BODE-) BODE GENE DEV CO LTD SHANGHAI.  
 XX Mao Y, Xie Y;  
 XX WPI; 2002-742038/81.  
 DR N-PSDB; ABZ25755.  
 XX Polypeptide-human protein 12.63 containing site-specific recombinase  
 PT characteristic sequence fragment and polynucleotide for coding it -



XX Claim 1, Page 27 (disclosure); 33pp; Chinese.

XX The invention relates to human site-specific recombinase motif-containing

CC protein 12.63 (ABP58560) and nucleic acids encoding it (AB25755). The

CC protein has a molecular weight of 12.63 kD. The invention also relates to

CC a method for the recombinant production of the protein, an antagonist of

CC the protein, and the use of the protein, gene and antagonist in

CC therapeutic applications. Site-specific recombinase motif-containing

CC protein 12.63 can be used in the treatment of a variety of diseases such

CC as cancer and HIV (human immunodeficiency virus) infection. The present

CC sequence represents human site-specific recombinase motif-containing

CC protein 12.63.

SQ Sequence 115 AA;

Query Match 87.8%; Score 115; DB 23; Length 115;

Best Local Similarity 100.0%; Pred. No. 7.7e-105; Indels 0; Gaps 0;

Matches 115; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 17 MSGEPQTSVAPPEVEPSSGVRIWVEYCEPCGFATYLELASAVKEQYPGIEISRLG 76

Db 1 MSGEPQTSVAPPEVEPSSGVRIWVEYCEPCGFATYLELASAVKEQYPGIEISRLG 60

QY 77 GTGAFIEINGQLVFSKLENGGFPYERKDLIEAIRRASNGETLEKINSRPPCVIL 131

Db 61 GTGAFIEINGQLVFSKLENGGFPYERKDLIEAIRRASNGETLEKINSRPPCVIL 115

RESULT 7

ABP43843

ID ABP43843 standard; Protein; 115 AA.

XX ABP43843;

AC

XX 26-FEB-2003 (first entry)

DT

XX RIKEN 1810046J19 protein.

DE

XX Neuroprotective; immunomodulator; cancer; chromosome 4q13-q21;

KM cytostatic; anti-inflammatory; gene therapy; nutritional supplement;

KW wound; burn; ulcer; Alzheimer's disease; Huntington's disease;

KW amyotrophic lateral sclerosis; autoimmune disorder; inflammation;

KW vlnetraty.

XX Homo sapiens.

OS

XX MO200231111-A2.

PN

XX 18-APR-2002.

PD

XX 11-OCT-2001; 2001MO-US27760.

PF

XX 12-OCT-2000; 2000US-0687527.

PR

XX (HYSE-) HYSEQ INC.

PA

XX Tang YT, Liu C, Zhou P, Aundt V, Zhang J, Zhao QA, Ren F;

PI Xue AJ, Yang Y, Wehrman T, Dmanac RT;

PI Xue AJ, Yang Y, Wehrman T, Dmanac RT;

PT WPI; 2002-426278/45.

PT N-PSDB; ABC61087.

DR

XX New polypeptides and their encoded proteins, useful as nutritional

PT sources or supplements, or in gene therapy, particularly for treating

PT wounds, Alzheimer's disease, amyotrophic lateral sclerosis, cancer or

PT inflammation -

XX Claim 20; SEQ ID # 746; 357pp + sequence listing; English.

PS The invention relates to 446 newly isolated polynucleotide sequences.

CC The activity of polynucleotides of the invention may be described as,

CC vulnerable, neuroprotective, immunomodulator, cytoskeletal and

CC anti-inflammatory. Compositions comprising nucleic acids of the invention

CC are useful for treating a mammalian subject, or as nutritional sources or

CC supplements. These are useful in gene therapy, particularly for treating

CC wounds, burns or ulcers, Alzheimer's disease, Huntington's disease,

CC amyotrophic lateral sclerosis, autoimmune disorders, cancer or

CC inflammation. The nucleic acids and polypeptides are also useful in

CC diagnostic and research methods. The sequences given in records

CC ABP43544-ABP43589 represent polypeptides encoded by polynucleotides of

CC the invention.

CC NOTE: The sequence data for this patent did not form part of the printed

CC specification, but was obtained in electronic format directly from WIPO

CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX

SQ Sequence 115 AA;

Query Match 87.8%; Score 115; DB 23; Length 115;

Best Local Similarity 100.0%; Pred. No. 7.7e-105; Indels 0; Gaps 0;

Matches 115; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 17 MSGEPQTSVAPPEVEPSSGVRIWVEYCEPCGFATYLELASAVKEQYPGIEISRLG 76

Db 1 MSGEPQTSVAPPEVEPSSGVRIWVEYCEPCGFATYLELASAVKEQYPGIEISRLG 60

QY 77 GTGAFIEINGQLVFSKLENGGFPYERKDLIEAIRRASNGETLEKINSRPPCVIL 131

Db 61 GTGAFIEINGQLVFSKLENGGFPYERKDLIEAIRRASNGETLEKINSRPPCVIL 115

RESULT 8

AA03153

ID AA03153 standard; Protein; 90 AA.

XX AA03153;

AC

XX 06-OCT-2000 (first entry)

DT

XX Human secreted protein, SEQ ID NO: 7234.

DE

XX Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;

KW gene therapy; chromosome mapping.

KW

XX Homo sapiens.

OS

XX EP1033401-A2.

PN

XX 06-SEP-2000.

PD

XX 21-FEB-2000; 2000EP-0200610.

PF

XX 26-FEB-1999; 99US-0122487.

PR

XX (GEST) GENSET.

PA

XX Dumas Mline Edwards J, Duclert A, Giordano J;

PI

XX WPI; 2000-500381/45.

PI N-PSDB; AAC03159.

DR

XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for

PT obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for

PT diagnostic, forensic, gene therapy and chromosome mapping procedures -

XX Claim 13; SEQ ID 7234; 71pp + CD-ROM; English.

PS The present sequence is a polypeptide encoded by one of a large number

CC of 5' ESTs derived from mRNAs encoding secreted proteins. The 5' ESTs

CC were prepared from total human RNAs or polyA+ RNAs derived from 30

CC different tissues. EST sequences usually correspond mainly to the 3'

CC untranslated region (UTR) of the mRNA because they are often obtained

CC from oligo-dT primed cDNA libraries. Such ESTs are not well suited for

CC isolating cDNA sequences derived from the 5' ends of mRNAs and even in

CC those cases where longer cDNA sequences have been obtained, the full 5'

CC UTR is rarely included. 5' ESTs are derived from mRNAs with intact 5'

CC ends and can therefore be used to obtain full length cDNAs and genomic  
CC DNAs. 5' ESTs are also used in diagnostic, forensic, gene therapy and  
CC chromosome mapping procedures. They are used to obtain upstream  
CC regulatory sequences and to design expression and secretion vectors.

XX Sequence 90 AA;

Query Match 20.6%; Score 27; DB 21; Length 90;  
Best Local Similarity 100.0%; Pred. No. 1.3e-18;  
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 35 PGSGVRIVEYCEPCGEATYLELASA 61  
|||  
DB 19 PGSGVRIVEYCEPCGEATYLELASA 45

RESULT 9  
AAG77898  
ID AAG77898 standard; Peptide; 21 AA.  
XX  
AC AAG77898;  
XX  
DT 08-MAY-2002 (first entry)  
XX

Human C35 peptide (amino acids 63-83), predicted HLA class II epitope.

XX C35 peptide; major histocompatibility complex; HLA class II epitope;  
XX MHC-peptide complex; MHC; human; MHC class I alpha chain;  
XX beta-2 microglobulin; MHC class II alpha chain; MHC class II beta chain;  
XX vaccine; immune response modulation; hyperproliferative disorder;  
XX neoplasm; hypergamma globulinaemia; viral infection; hepatitis;  
XX meningitis; bacterial infection; tuberculosis; gingivitis;  
XX parasitic infection; autoimmune disease; Hashimoto's disease;  
XX Graves' disease; rheumatoid arthritis; allergy; asthma; organ rejection;  
XX graft-versus-host disease; GVHD; breast cancer; human leukocyte antigen.

XX Homo sapiens.

XX WO200178768-A2.

XX 25-OCT-2001.

XX 12-APR-2001; 2001WO-US11912.

XX 12-APR-2000; 2000US-196472P.

XX (UTRP) UNIV ROCHESTER.

XX Zauderer M, Smith ES;

XX WPI; 2001-602927/68.

XX Novel compound comprising major histocompatibility complex-peptide

XX complexes, used to modulate immune responses -  
XX  
XX Disclosure; Page 19; 16pp; English.

XX The invention comprises a compound which contains one or more major  
XX histocompatibility complex (MHC)-peptide complexes, and an antibody  
XX specific for a cell surface marker. The complexes comprise an MHC class  
XX I alpha chain, a beta-2 microglobulin molecule and an antigenic peptide  
XX bound in the MHC groove. Alternatively, the complexes may comprise an MHC  
XX class II alpha chain, an MHC class II beta chain, and an antigenic  
XX peptide bound in the MHC groove. The complexes are linked to the carboxyl  
XX terminus of the antibody. The compounds of the invention can be used as a  
XX vaccine to modulate an immune response. The compounds of the invention  
XX are useful for treating: hyperproliferative disorders (e.g. neoplasms and  
XX hypergamma globulinaemia); viral infections (e.g. hepatitis and  
XX meningitis); bacterial infections (e.g. tuberculosis and gingivitis);  
XX parasitic infections; autoimmune diseases (e.g. Hashimoto's disease,  
XX Graves' disease and rheumatoid arthritis); allergic reactions/conditions  
XX (e.g. asthma). The compounds of the invention may also be used in the  
XX treatment of organ rejection or graft-versus-host disease (GVHD). The

CC present sequence represents amino acids 63-83 of the human C35 protein,  
CC which is differentially expressed in breast cancer cells. The present  
CC sequence is a predicted C35 human leukocyte antigen (HLA) class II  
CC epitope.

XX Sequence 21 AA;

Query Match 16.0%; Score 21; DB 22; Length 21;  
Best Local Similarity 100.0%; Pred. No. 2.7e-13;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 79 GAFEIRINGLVFSKLENGGF 99  
|||  
DB 1 GAFEIRINGLVFSKLENGGF 21

RESULT 10  
AAG77899  
ID AAG77899 standard; Peptide; 21 AA.  
XX  
AC AAG77899;  
XX  
DT 08-MAY-2002 (first entry)  
XX

Human C35 peptide (amino acids 83-103), predicted HLA class II epitope.

XX C35 peptide; major histocompatibility complex; HLA class II epitope;  
XX MHC-peptide complex; MHC; human; MHC class I alpha chain;  
XX beta-2 microglobulin; MHC class II alpha chain; MHC class II beta chain;  
XX vaccine; immune response modulation; hyperproliferative disorder;  
XX neoplasm; hypergamma globulinaemia; viral infection; hepatitis;  
XX meningitis; bacterial infection; tuberculosis; gingivitis;  
XX parasitic infection; autoimmune disease; Hashimoto's disease;  
XX Graves' disease; rheumatoid arthritis; allergy; asthma; organ rejection;  
XX graft-versus-host disease; GVHD; breast cancer; human leukocyte antigen.

XX Homo sapiens.

XX WO200178768-A2.

XX 25-OCT-2001.

XX 12-APR-2001; 2001WO-US11912.

XX 12-APR-2000; 2000US-196472P.

XX (UTRP) UNIV ROCHESTER.

XX Zauderer M, Smith ES;

XX WPI; 2001-602927/68.

XX Novel compound comprising major histocompatibility complex-peptide

XX complexes, used to modulate immune responses -  
XX  
XX Disclosure; Page 19; 16pp; English.

XX The invention comprises a compound which contains one or more major  
XX histocompatibility complex (MHC)-peptide complexes, and an antibody  
XX specific for a cell surface marker. The complexes comprise an MHC class  
XX I alpha chain, a beta-2 microglobulin molecule and an antigenic peptide  
XX bound in the MHC groove. Alternatively, the complexes may comprise an MHC  
XX class II alpha chain, an MHC class II beta chain, and an antigenic  
XX peptide bound in the MHC groove. The complexes are linked to the carboxyl  
XX terminus of the antibody. The compounds of the invention can be used as a  
XX vaccine to modulate an immune response. The compounds of the invention  
XX are useful for treating: hyperproliferative disorders (e.g. neoplasms and  
XX hypergamma globulinaemia); viral infections (e.g. hepatitis and  
XX meningitis); bacterial infections (e.g. tuberculosis and gingivitis);  
XX parasitic infections; autoimmune diseases (e.g. Hashimoto's disease,  
XX Graves' disease and rheumatoid arthritis); allergic reactions/conditions  
XX (e.g. asthma). The compounds of the invention may also be used in the  
XX treatment of organ rejection or graft-versus-host disease (GVHD). The

CC present sequence represents amino acids 83-103 of the human C35 protein,  
 CC which is differentially expressed in breast cancer cells. The present  
 CC sequence is a predicted C35 human leukocyte antigen (HLA) class II  
 CC epitope.

XX  
 XX  
 SQ Sequence 21 AA;

Query Match 16.0%; Score 21; DB 22; Length 21;  
 Best Local Similarity 100.0%; Pred. No. 2.7e-13;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 99 FPEKDLIEAIRRASNGETLE 119  
 |||||  
 DB 1 FPEKDLIEAIRRASNGETLE 21

# RESULT 11

ABBI4673  
 ID ABBI4673 standard; Peptide; 21 AA.

XX  
 AC ABBI4673;

XX  
 DT 22-JAN-2002 (first entry)

XX  
 DE Human C35 peptide epitope #2177.

XX  
 KW Human; C35; cytostatic; gene therapy; vaccine; tumour antigen;  
 KW breast cancer; bladder cancer; tumour immunotherapy; epitope;  
 KW major histocompatibility complex binding peptide; MHC.

XX  
 OS Homo sapiens.

XX  
 PN WO200174859-A2.

XX  
 PD 11-OCT-2001.

XX  
 PF 04-APR-2001; 2001WO-US10855.

XX  
 PR 04-APR-2000; 2000US-19463P.

XX  
 PA (UYRP ) UNIV ROCHESTER.

XX  
 PI Zauderer M, Evans EE, Borrello MA;

XX  
 DR WPI; 2001-626383/72.

XX  
 PT Novel C35 polypeptides and C35 genes useful in immunogenic compositions  
 PT and vaccines, for inducing antibody and cell-mediated immunity against  
 PT target cells, such as tumor cells that express C35 gene -

XX  
 PS Disclosure; Page 178; 331pp; English.

XX  
 CC The present invention relates to human C35 (see AAG78997). C35 is a novel  
 CC tumour antigen that is overexpressed in human breast and bladder  
 CC carcinoma. C35 is thought to be a promising candidate for tumour  
 CC immunotherapy, in immunogenic compositions and vaccines, to induce  
 CC antibody and cell-mediated immunity against target cells such as tumour  
 CC cells that express C35 genes. The present sequence is a C35 peptide  
 CC epitope. This peptide is predicted to be a major histocompatibility  
 CC complex (MHC) binding peptide.

XX  
 SQ Sequence 21 AA;

Query Match 16.0%; Score 21; DB 22; Length 21;  
 Best Local Similarity 100.0%; Pred. No. 2.7e-13;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 79 GAFERINQGVLSKLENGGF 99  
 |||||  
 DB 1 GAFERINQGVLSKLENGGF 21

# RESULT 12

ABBI4674  
 ID ABBI4674 standard; Peptide; 21 AA.

XX  
 AC ABBI4674;

XX  
 DT 22-JAN-2002 (first entry)

XX  
 DE Human C35 peptide epitope #2178.

XX  
 KW Human; C35; cytostatic; gene therapy; vaccine; tumour antigen;  
 KW breast cancer; bladder cancer; tumour immunotherapy; epitope;  
 KW major histocompatibility complex binding peptide; MHC.

XX  
 OS Homo sapiens.

XX  
 PN WO200174859-A2.

XX  
 PD 11-OCT-2001.

XX  
 PF 04-APR-2001; 2001WO-US10855.

XX  
 PR 04-APR-2000; 2000US-19463P.

XX  
 PA (UYRP ) UNIV ROCHESTER.

XX  
 PI Zauderer M, Evans EE, Borrello MA;

XX  
 DR WPI; 2001-626383/72.

XX  
 PT Novel C35 polypeptides and C35 genes useful in immunogenic compositions  
 PT and vaccines, for inducing antibody and cell-mediated immunity against  
 PT target cells, such as tumor cells that express C35 gene -

XX  
 PS Disclosure; Page 178; 331pp; English.

XX  
 CC The present invention relates to human C35 (see AAG78997). C35 is a novel  
 CC tumour antigen that is overexpressed in human breast and bladder  
 CC carcinoma. C35 is thought to be a promising candidate for tumour  
 CC immunotherapy, in immunogenic compositions and vaccines, to induce  
 CC antibody and cell-mediated immunity against target cells such as tumour  
 CC cells that express C35 genes. The present sequence is a C35 peptide  
 CC epitope. This peptide is predicted to be a major histocompatibility  
 CC complex (MHC) binding peptide.

XX  
 SQ Sequence 21 AA;

Query Match 16.0%; Score 21; DB 22; Length 21;  
 Best Local Similarity 100.0%; Pred. No. 2.7e-13;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 99 FPEKDLIEAIRRASNGETLE 119  
 |||||  
 DB 1 FPEKDLIEAIRRASNGETLE 21

# RESULT 13

AAG7901  
 ID AAG7901 standard; Peptide; 19 AA.

XX  
 AC AAG7901;

XX  
 DT 08-MAY-2002 (first entry)

XX  
 DE Human C35 peptide (amino acids 57-75), predicted HLA class II epitope.

XX  
 KW C35 peptide; major histocompatibility complex; HLA class II epitope;  
 KW MHC-peptide complex; MHC; human; MHC class I alpha chain;  
 KW beta-2 microglobulin; MHC class II alpha chain; MHC class II beta chain;  
 KW vaccine; immune response modulation; hyperproliferative disorder;  
 KW neoplasia; hyperimmunoglobulinemia; viral infection; hepatitis;  
 KW meningitis; bacterial infection; tuberculosis; gingivitis;  
 KW parasitic infection; autoimmune disease; Hashimoto's disease;  
 KW Graves' disease; rheumatoid arthritis; allergy; asthma; organ rejection;

KM graft-versus-host disease; GVHD; breast cancer; human leukocyte antigen.  
XX Homo sapiens.  
OS  
XX WO200178768-A2.  
XX  
XX 25-OCT-2001.  
XX  
XX 12-APR-2001; 2001WO-US11912.  
XX  
XX 12-APR-2000; 2000US-196472P.  
XX  
XX (UYRP ) UNIV ROCHESTER.  
XX  
XX Zauderer M, Smith ES;  
XX  
XX WPI; 2001-602927/68.  
XX  
XX  
XX Novel compound comprising major histocompatibility complex-peptide  
XX complexes, used to modulate immune responses -  
XX  
XX  
XX Disclosure; Page 19; 166pp; English.  
XX  
XX The invention comprises a compound which contains one or more major  
XX histocompatibility complex (MHC)-peptide complexes, and an antibody  
XX specific for a cell surface marker. The complexes comprise an MHC class  
XX I alpha chain, a beta-2 microglobulin molecule and an antigenic peptide  
XX bound in the MHC groove. Alternatively, the complexes may comprise an MHC  
XX class II alpha chain, an MHC class II beta chain, and an antigenic  
XX peptide bound in the MHC groove. The complexes are linked to the carboxyl  
XX terminus of the antibody. The compounds of the invention can be used as a  
XX vaccine to modulate an immune response. The compounds of the invention  
XX are useful for treating: hyperproliferative disorders (e.g. neoplasms and  
XX hypergammaglobulinemia); viral infections (e.g. hepatitis and  
XX meningitis); bacterial infections (e.g. tuberculosis and gingivitis);  
XX parasitic infections; autoimmune diseases (e.g. Hashimoto's disease,  
XX Graves' disease and rheumatoid arthritis); allergic reactions/conditions  
XX (e.g. asthma). The compounds of the invention may also be used in the  
XX treatment of organ rejection or graft-versus-host disease (GVHD). The  
XX present sequence represents amino acids 57-75 of the human C35 protein,  
XX which is differentially expressed in breast cancer cells. The present  
XX sequence is a predicted C35 human leukocyte antigen (HLA) class II  
XX epitope.  
XX  
XX Sequence 19 AA;  
SQ  
Query Match 14.5%; Score 19; DB 22; Length 19;  
Best Local Similarity 100.0%; Pred. No. 2.3e-11;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 73 SRLGGTGAFFEINGQLVF 91  
Db 1 SRLGGTGAFFEINGQLVF 19  
RESULT 14  
ABBI4672  
ID ABBI4672 standard; Peptide; 19 AA.  
XX  
XX ABB14672;  
XX  
XX 22-JAN-2002 (first entry)  
XX  
XX Human C35 peptide epitope #2176.  
XX  
XX Human; C35; cytostatic; gene therapy; vaccine; tumour antigen;  
XX breast cancer; bladder cancer; tumour immunotherapy; epitope;  
XX major histocompatibility complex binding peptide; MHC.  
XX  
XX Homo sapiens.  
XX  
XX WO200174859-A2.  
XX  
XX

PD 11-OCT-2001.  
XX  
XX 04-APR-2001; 2001WO-US10855.  
XX  
XX 04-APR-2000; 2000US-194463P.  
XX  
XX (UYRP ) UNIV ROCHESTER.  
XX  
XX Zauderer M, Evans EE, Borrello MA;  
XX  
XX WPI; 2001-626383/72.  
XX  
XX  
XX Novel C35 polypeptides and C35 genes useful in immunogenic compositions  
XX and vaccines, for inducing antibody and cell-mediated immunity against  
XX target cells, such as tumor cells that express C35 gene -  
XX  
XX  
XX Disclosure; Page 177; 331pp; English.  
XX  
XX The present invention relates to human C35 (see AAG78997). C35 is a novel  
XX tumour antigen that is overexpressed in human breast and bladder  
XX carcinoma. C35 is thought to be a promising candidate for tumour  
XX immunotherapy, in immunogenic compositions and vaccines, to induce  
XX antibody and cell-mediated immunity against target cells such as tumour  
XX cells that express C35 genes. The present sequence is a C35 peptide  
XX epitope. This peptide is predicted to be a major histocompatibility  
XX complex (MHC) binding peptide.  
XX  
XX Sequence 19 AA;  
SQ  
Query Match 14.5%; Score 19; DB 22; Length 19;  
Best Local Similarity 100.0%; Pred. No. 2.3e-11;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 73 SRLGGTGAFFEINGQLVF 91  
Db 1 SRLGGTGAFFEINGQLVF 19  
RESULT 15  
AAG77900  
ID AAG77900 standard; Peptide; 15 AA.  
XX  
XX AAG77900;  
XX  
XX 08-MAY-2002 (first entry)  
XX  
XX  
XX Human C35 peptide (amino acids 21-35), predicted HLA class II epitope.  
XX  
XX C35 peptide; major histocompatibility complex; HLA class II epitope;  
XX MHC-peptide complex; MHC; human; MHC class I alpha chain;  
XX beta-2 microglobulin; MHC class II alpha chain; MHC class II beta chain;  
XX vaccine; immune response modulation; hyperproliferative disorder;  
XX neoplasms; hypergammaglobulinemia; viral infection; hepatitis;  
XX meningitis; bacterial infection; tuberculosis; gingivitis;  
XX parasitic infection; autoimmune disease; Hashimoto's disease;  
XX Graves' disease; rheumatoid arthritis; allergy; asthma; organ rejection;  
XX graft-versus-host disease; GVHD; breast cancer; human leukocyte antigen.  
XX  
XX Homo sapiens.  
XX  
XX WO200178768-A2.  
XX  
XX 25-OCT-2001.  
XX  
XX 12-APR-2001; 2001WO-US11912.  
XX  
XX 12-APR-2000; 2000US-196472P.  
XX  
XX (UYRP ) UNIV ROCHESTER.  
XX  
XX Zauderer M, Smith ES;  
XX  
XX WPI; 2001-602927/68.  
XX  
XX

XX Novel compound comprising major histocompatibility complex-peptide  
PT complexes, used to modulate immune responses -  
XX Disclosure; Page 19, 16pp; English.  
XX The invention comprises a compound which contains one or more major  
CC histocompatibility complex (MHC)-peptide complexes, and an antibody  
CC specific for a cell surface marker. The complexes comprise an MHC class  
CC I alpha chain, a beta-2-microglobulin molecule and an antigenic peptide  
CC bound in the MHC groove. Alternatively, the complexes may comprise an MHC  
CC class II alpha chain, an MHC class II beta chain, and an antigenic  
CC peptide bound in the MHC groove. The complexes are linked to the carboxyl  
CC terminus of the antibody. The compounds of the invention can be used as a  
CC vaccine to modulate an immune response. The compounds of the invention  
CC are useful for treating: hyperproliferative disorders (e.g. neoplasms and  
CC hypermaglobulinemia); viral infections (e.g. hepatitis and  
CC meningitis); bacterial infections (e.g. tuberculosis and gingivitis);  
CC parasitic infections; autoimmune diseases (e.g. Hashimoto's disease,  
CC Graves' disease and rheumatoid arthritis); allergic reactions/conditions  
CC (e.g. asthma). The compounds of the invention may also be used in the  
CC treatment of organ rejection or graft-versus-host disease (GVHD). The  
CC present sequence represents amino acids 21-35 of the human C35 protein,  
CC which is differentially expressed in breast cancer cells. The present  
CC sequence is a predicted C35 human leukocyte antigen (HLA) class II  
CC epitope.  
SQ Sequence 15 AA;  
Query Match 11.5%; Score 15; DB 22; Length 15;  
Best Local Similarity 100.0%; Pred. No. 1.5e-07;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 37 SGVRIVYCEPCGF 51  
DB 1 SGVRIVYCEPCGF 15  
RESULT 16  
ABBI2825  
ID ABBI2825 standard; Peptide; 15 AA.  
AC ABBI2825;  
XX 22-JAN-2002 (first entry)  
XX Human C35 peptide epitope #329.  
DE Human C35 peptide epitope #329.  
XX Human; C35; cytostatic; gene therapy; vaccine; tumour antigen;  
KW breast cancer; bladder cancer; tumour immunotherapy; epitope;  
KM major histocompatibility complex binding peptide; MHC.  
XX Homo sapiens.  
OS WO200174859-A2.  
PN 11-OCT-2001.  
PD 04-APR-2001; 2001WO-US10855.  
XX 04-APR-2000; 2000US-194463P.  
PR (UVRP ) UNIV ROCHESTER.  
PA Zauderer M, Evans EE, Borrello MA,  
XX WPI, 2001-626383/72.  
DR Novel C35 polypeptides and C35 genes useful in immunogenic compositions  
XX PT and vaccines, for inducing antibody and cell-mediated immunity against  
XX target cells, such as tumor cells that express C35 gene -  
XX Disclosure; Page 60; 331pp; English.

XX The present invention relates to human C35 (see AAG78997). C35 is a novel  
CC tumour antigen that is overexpressed in human breast and bladder  
CC carcinoma. C35 is thought to be a promising candidate for tumour  
CC immunotherapy, in immunogenic compositions and vaccines, to induce  
CC antibody and cell-mediated immunity against target cells such as tumour  
CC cells that express C35 genes. The present sequence is a C35 peptide  
CC epitope. This peptide is predicted to be a major histocompatibility  
CC complex (MHC) binding peptide.  
SQ Sequence 15 AA;  
Query Match 11.5%; Score 15; DB 22; Length 15;  
Best Local Similarity 100.0%; Pred. No. 1.5e-07;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 88 QLVFSLKENGFPYE 102  
DB 1 QLVFSLKENGFPYE 15  
RESULT 17  
ABBI2826  
ID ABBI2826 standard; Peptide; 15 AA.  
AC ABBI2826;  
XX 22-JAN-2002 (first entry)  
XX Human C35 peptide epitope #330.  
DE Human C35 peptide epitope #330.  
XX Human; C35; cytostatic; gene therapy; vaccine; tumour antigen;  
KW breast cancer; bladder cancer; tumour immunotherapy; epitope;  
KM major histocompatibility complex binding peptide; MHC.  
XX Homo sapiens.  
OS WO200174859-A2.  
PN 11-OCT-2001.  
PD 04-APR-2001; 2001WO-US10855.  
XX 04-APR-2000; 2000US-194463P.  
PR (UVRP ) UNIV ROCHESTER.  
PA Zauderer M, Evans EE, Borrello MA,  
XX WPI, 2001-626383/72.  
DR Novel C35 polypeptides and C35 genes useful in immunogenic compositions  
XX PT and vaccines, for inducing antibody and cell-mediated immunity against  
XX target cells, such as tumor cells that express C35 gene -  
XX Disclosure; Page 60; 331pp; English.  
XX The present invention relates to human C35 (see AAG78997). C35 is a novel  
CC tumour antigen that is overexpressed in human breast and bladder  
CC carcinoma. C35 is thought to be a promising candidate for tumour  
CC immunotherapy, in immunogenic compositions and vaccines, to induce  
CC antibody and cell-mediated immunity against target cells such as tumour  
CC cells that express C35 genes. The present sequence is a C35 peptide  
CC epitope. This peptide is predicted to be a major histocompatibility  
CC complex (MHC) binding peptide.  
SQ Sequence 15 AA;  
Query Match 11.5%; Score 15; DB 22; Length 15;  
Best Local Similarity 100.0%; Pred. No. 1.5e-07;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 53 ATYELIASVKEQYP 67

Db 1 ATYLEILASAVEQYP 15

## RESULT 18

ABBI2827 standard; Peptide; 15 AA.

AC ABBI2827;

DT 22-JAN-2002 (first entry)

DE Human C35 peptide epitope #331.

XX Human; C35; cytostatic; gene therapy; vaccine; tumour antigen;

KM breast cancer; bladder cancer; tumour immunotherapy; epitope;

XX major histocompatibility complex binding peptide; MHC.

OS Homo sapiens.

XX WO200174859-A2.

XX 11-OCT-2001.

XX 04-APR-2001; 2001WO-US10855.

XX 04-APR-2000; 2000US-194463P.

XX (UTRP ) UNIV ROCHESTER.

XX Zauderer M, Evans EE, Borrello MA;

XX WPI; 2001-626383/72.

XX Novel C35 polypeptides and C35 genes useful in immunogenic compositions

XX and vaccines, for inducing antibody and cell-mediated immunity against

XX target cells, such as tumor cells that express C35 gene

XX Disclosure; Page 60; 31pp; English.

XX The present invention relates to human C35 (see AAG78997). C35 is a novel

XX tumour antigen that is overexpressed in human breast and bladder

XX carcinoma. C35 is thought to be a promising candidate for tumour

XX immunotherapy, in immunogenic compositions and vaccines, to induce

XX antibody and cell-mediated immunity against target cells such as tumour

XX cells that express C35 genes. The present sequence is a C35 peptide

XX epitope. This peptide is predicted to be a major histocompatibility

XX complex (MHC) binding peptide.

## RESULT 19

ABBI2828 standard; Peptide; 15 AA.

XX ABBI2828;

XX 22-JAN-2002 (first entry)

XX Human C35 peptide epitope #332.

XX Human; C35; cytostatic; gene therapy; vaccine; tumour antigen;

KM breast cancer; bladder cancer; tumour immunotherapy; epitope;

XX major histocompatibility complex binding peptide; MHC.

XX Homo sapiens.

XX WO200174859-A2.

XX 11-OCT-2001.

XX 04-APR-2001; 2001WO-US10855.

XX 04-APR-2000; 2000US-194463P.

XX (UTRP ) UNIV ROCHESTER.

XX Zauderer M, Evans EE, Borrello MA;

XX WPI; 2001-626383/72.

XX Novel C35 polypeptides and C35 genes useful in immunogenic compositions

XX and vaccines, for inducing antibody and cell-mediated immunity against

XX target cells, such as tumor cells that express C35 gene

XX Disclosure; Page 60; 31pp; English.

XX The present invention relates to human C35 (see AAG78997). C35 is a novel

XX tumour antigen that is overexpressed in human breast and bladder

XX carcinoma. C35 is thought to be a promising candidate for tumour

XX immunotherapy, in immunogenic compositions and vaccines, to induce

XX antibody and cell-mediated immunity against target cells such as tumour

XX cells that express C35 genes. The present sequence is a C35 peptide

XX epitope. This peptide is predicted to be a major histocompatibility

XX complex (MHC) binding peptide.

XX Sequence 15 AA;

XX Query March 11.5%; Score 15; DB 22; Length 15;

XX Best Local Similarity 100.0%; Pred. No. 1.5e-07;

XX Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

XX RESULT 20

XX ABBI2829 standard; Peptide; 15 AA.

XX ABBI2829;

XX 22-JAN-2002 (first entry)

XX Human C35 peptide epitope #333.

XX Human; C35; cytostatic; gene therapy; vaccine; tumour antigen;

KM breast cancer; bladder cancer; tumour immunotherapy; epitope;

XX major histocompatibility complex binding peptide; MHC.

XX Homo sapiens.

XX WO200174859-A2.

XX 11-OCT-2001.

XX 04-APR-2001; 2001WO-US10855.

XX 04-APR-2000; 2000US-194463P.

XX (UTRP ) UNIV ROCHESTER.

XX Zauderer M, Evans EE, Borrello MA;

XX WPI; 2001-626383/72.

PT Novel C35 polypeptides and C35 genes useful in immunogenic compositions  
PT and vaccine, for inducing antibody and cell-mediated immunity against  
PT target cells, such as tumor cells that express C35 gene -  
XX  
XX  
PS Disclosure; Page 60; 331pp; English.  
XX  
CC The present invention relates to human C35 (see AAG78997). C35 is a novel  
CC tumour antigen that is overexpressed in human breast and bladder  
CC carcinoma. C35 is thought to be a promising candidate for tumour  
CC immunotherapy, in immunogenic compositions and vaccines, to induce  
CC antibody and cell-mediated immunity against target cells such as tumour  
CC cells that express C35 genes. The present sequence is a C35 peptide  
CC epitope. This peptide is predicted to be a major histocompatibility  
CC complex (MHC) binding peptide.  
XX  
SQ Sequence 15 AA;  
  
Query Match 11.5%; Score 15; DB 22; Length 15;  
Best Local Similarity 100.0%; Pred. No. 1.5e-07;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 40 RIVVCEPCGFEAT 54  
Db 1 RIVVCEPCGFEAT 15  
  
RESULT 21  
ABBI2830  
ID ABBI2830 standard; Peptide; 15 AA.  
AC ABBI2830;  
XX  
XX 22-JAN-2002 (first entry)  
DT  
XX  
DE Human C35 peptide epitope #334.  
XX  
XX Human; C35; cytostatic; gene therapy; vaccine; tumour antigen;  
XX breast cancer; bladder cancer; tumour immunotherapy; epitope;  
XX major histocompatibility complex binding peptide; MHC.  
XX  
XX Homo sapiens.  
XX  
XX WO200174859-A2.  
XX  
XX 11-OCT-2001.  
XX  
XX 04-APR-2001; 2001WO-US10855.  
XX  
XX 04-APR-2000; 2000US-194463P.  
XX  
XX (UYRP ) UNITV ROCHESTER.  
XX  
XX Zauderer M, Evans EE, Borrello MA;  
XX  
XX WPI; 2001-626383/72.  
XX  
XX  
XX Novel C35 polypeptides and C35 genes useful in immunogenic compositions  
XX and vaccine, for inducing antibody and cell-mediated immunity against  
XX target cells, such as tumor cells that express C35 gene -  
XX  
XX  
XX Disclosure; Page 60; 331pp; English.  
XX  
XX The present invention relates to human C35 (see AAG78997). C35 is a novel  
XX tumour antigen that is overexpressed in human breast and bladder  
XX carcinoma. C35 is thought to be a promising candidate for tumour  
XX immunotherapy, in immunogenic compositions and vaccines, to induce  
XX antibody and cell-mediated immunity against target cells such as tumour  
XX cells that express C35 genes. The present sequence is a C35 peptide  
XX epitope. This peptide is predicted to be a major histocompatibility  
XX complex (MHC) binding peptide.  
XX  
XX Sequence 15 AA;  
  
Query Match 11.5%; Score 15; DB 22; Length 15;  
Best Local Similarity 100.0%; Pred. No. 1.5e-07;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 55 YIELASAVKEQYPGI 69  
Db 1 YIELASAVKEQYPGI 15  
  
RESULT 23  
ABBI2832  
ID ABBI2832 standard; Peptide; 15 AA.  
AC ABBI2832;  
XX  
XX 22-JAN-2002 (first entry)  
DT  
XX

Query Match 11.5%; Score 15; DB 22; Length 15;  
Best Local Similarity 100.0%; Pred. No. 1.5e-07;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 52 EATYLELASAVKEQY 66  
Db 1 EATYLELASAVKEQY 15  
  
RESULT 22  
ABBI2831  
ID ABBI2831 standard; Peptide; 15 AA.  
AC ABBI2831;  
XX  
XX 22-JAN-2002 (first entry)  
DT  
XX  
DE Human C35 peptide epitope #335.  
XX  
XX Human; C35; cytostatic; gene therapy; vaccine; tumour antigen;  
XX breast cancer; bladder cancer; tumour immunotherapy; epitope;  
XX major histocompatibility complex binding peptide; MHC.  
XX  
XX Homo sapiens.  
XX  
XX WO200174859-A2.  
XX  
XX 11-OCT-2001.  
XX  
XX 04-APR-2001; 2001WO-US10855.  
XX  
XX 04-APR-2000; 2000US-194463P.  
XX  
XX (UYRP ) UNITV ROCHESTER.  
XX  
XX Zauderer M, Evans EE, Borrello MA;  
XX  
XX WPI; 2001-626383/72.  
XX  
XX  
XX Novel C35 polypeptides and C35 genes useful in immunogenic compositions  
XX and vaccine, for inducing antibody and cell-mediated immunity against  
XX target cells, such as tumor cells that express C35 gene -  
XX  
XX  
XX Disclosure; Page 60; 331pp; English.  
XX  
XX The present invention relates to human C35 (see AAG78997). C35 is a novel  
XX tumour antigen that is overexpressed in human breast and bladder  
XX carcinoma. C35 is thought to be a promising candidate for tumour  
XX immunotherapy, in immunogenic compositions and vaccines, to induce  
XX antibody and cell-mediated immunity against target cells such as tumour  
XX cells that express C35 genes. The present sequence is a C35 peptide  
XX epitope. This peptide is predicted to be a major histocompatibility  
XX complex (MHC) binding peptide.  
XX  
XX Sequence 15 AA;  
  
Query Match 11.5%; Score 15; DB 22; Length 15;  
Best Local Similarity 100.0%; Pred. No. 1.5e-07;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 55 YIELASAVKEQYPGI 69  
Db 1 YIELASAVKEQYPGI 15  
  
RESULT 23  
ABBI2832  
ID ABBI2832 standard; Peptide; 15 AA.  
AC ABBI2832;  
XX  
XX 22-JAN-2002 (first entry)  
DT  
XX

XX Zauderer M, Evans EE, Borrello MA;  
 PI  
 XX WP1; 2001-626383/72.  
 DR  
 XX Novel C35 polypeptides and C35 genes useful in immunogenic compositions  
 PT and vaccines, for inducing antibody and cell-mediated immunity against  
 PT target cells, such as tumor cells that express C35 gene -  
 XX  
 PS Disclosure; Page 60; 331pp; English.  
 XX  
 CC The present invention relates to human C35 (see AAG78997). C35 is a novel  
 CC tumour antigen that is overexpressed in human breast and bladder  
 CC carcinoma. C35 is thought to be a promising candidate for tumour  
 CC immunotherapy, in immunogenic compositions and vaccines, to induce  
 CC antibody and cell-mediated immunity against target cells such as tumour  
 CC cells that express C35 genes. The present sequence is a C35 peptide  
 CC epitope. This peptide is predicted to be a major histocompatibility  
 CC complex (MHC) binding peptide.  
 XX  
 SQ Sequence 15 AA;  
 Query Match 11.5%; Score 15; DB 22; Length 15;  
 Best Local Similarity 100.0%; Pred. No. 1.5e-07;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Oy 72 ESRIGTGAFRIEIN 86  
 |||||  
 Db 1 ESRIGTGAFRIEIN 15  
 RESULT 25  
 ABB12834  
 ID ABB12834 standard; Peptide; 15 AA.  
 AC ABB12834;  
 XX  
 DT 22-JAN-2002 (first entry)  
 DE Human C35 peptide epitope #338.  
 XX  
 XX Human; C35; cytostatic; gene therapy; vaccine; tumour antigen;  
 KW breast cancer; bladder cancer; tumour immunotherapy; epitope;  
 KW major histocompatibility complex binding peptide; MHC.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200174859-A2.  
 PD 11-OCT-2001.  
 PF 04-APR-2001; 2001WO-US10855.  
 XX  
 PR 04-APR-2000; 2000US-194463P.  
 PA (UTRP ) UNIV ROCHESTER.  
 XX  
 PI Zauderer M, Evans EE, Borrello MA;  
 DR WP1; 2001-626383/72.  
 XX  
 PT Novel C35 polypeptides and C35 genes useful in immunogenic compositions  
 PT and vaccines, for inducing antibody and cell-mediated immunity against  
 PT target cells, such as tumor cells that express C35 gene -  
 XX  
 PS Disclosure; Page 60; 331pp; English.  
 XX  
 CC The present invention relates to human C35 (see AAG78997). C35 is a novel  
 CC tumour antigen that is overexpressed in human breast and bladder  
 CC carcinoma. C35 is thought to be a promising candidate for tumour  
 CC immunotherapy, in immunogenic compositions and vaccines, to induce  
 CC antibody and cell-mediated immunity against target cells such as tumour  
 CC cells that express C35 genes. The present sequence is a C35 peptide  
 CC cells that express C35 genes. The present sequence is a C35 peptide



CC epitope. This peptide is predicted to be a major histocompatibility  
 CC complex (MHC) binding peptide.

XX  
 XX  
 SQ Sequence 15 AA;

Query Match 11.5%; Score 15; DB 22; Length 15;

Best Local Similarity 100.0%; Pred. No. 1.5e-07;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 30 PEEVPSGVRIVE 44  
 |||||  
 1 PEEVPSGVRIVE 15

RESULT 26  
 ABB12835  
 ID ABB12835 standard; Peptide; 15 AA.

XX  
 XX ABB12835;

XX  
 XX 22-JAN-2002 (first entry)

XX  
 XX Human C35 peptide epitope #339.

KM Human; C35; cytostatic; gene therapy; vaccine; tumour antigen;  
 KM breast cancer; bladder cancer; tumour immunotherapy; epitope;  
 KM major histocompatibility complex binding peptide; MHC.

XX  
 XX Homo sapiens.

XX  
 XX WO200174859-A2.

XX  
 XX 11-OCT-2001.

XX  
 XX 04-APR-2001; 2001WO-US10855.

XX  
 XX 04-APR-2000; 2000US-19463P.

XX  
 XX (UYRP ) UNIV ROCHESTER.

XX  
 XX Zauderer M, Evans EE, Borrello MA;

XX  
 XX WPI; 2001-626383/72.

XX  
 XX Novel C35 polypeptides and C35 genes useful in immunogenic compositions  
 PT and vaccines, for inducing antibody and cell-mediated immunity against  
 PT target cells, such as tumor cells that express C35 gene -

XX  
 XX Disclosure; Page 61; 331pp; English.

XX  
 XX The present invention relates to human C35 (see AAG78997). C35 is a novel  
 CC tumour antigen that is overexpressed in human breast and bladder  
 CC carcinoma. C35 is thought to be a promising candidate for tumour  
 CC immunotherapy, in immunogenic compositions and vaccines, to induce  
 CC antibody and cell-mediated immunity against target cells such as tumour  
 CC cells that express C35 genes. The present sequence is a C35 peptide  
 CC epitope. This peptide is predicted to be a major histocompatibility  
 CC complex (MHC) binding peptide.

XX  
 XX Sequence 15 AA;

Query Match 11.5%; Score 15; DB 22; Length 15;

Best Local Similarity 100.0%; Pred. No. 1.5e-07;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 59 ASAVKEQYPGIETES 73  
 |||||  
 1 ASAVKEQYPGIETES 15

RESULT 27

ABB12836  
 ID ABB12836 standard; Peptide; 15 AA.

XX  
 AC ABB12836;

XX  
 XX 22-JAN-2002 (first entry)

XX  
 XX Human C35 peptide epitope #340.

KM Human; C35; cytostatic; gene therapy; vaccine; tumour antigen;  
 KM breast cancer; bladder cancer; tumour immunotherapy; epitope;  
 KM major histocompatibility complex binding peptide; MHC.

XX  
 XX Homo sapiens.

XX  
 XX WO200174859-A2.

XX  
 XX 11-OCT-2001.

XX  
 XX 04-APR-2001; 2001WO-US10855.

XX  
 XX 04-APR-2000; 2000US-19463P.

XX  
 XX (UYRP ) UNIV ROCHESTER.

XX  
 XX Zauderer M, Evans EE, Borrello MA;

XX  
 XX WPI; 2001-626383/72.

XX  
 XX Novel C35 polypeptides and C35 genes useful in immunogenic compositions  
 PT and vaccines, for inducing antibody and cell-mediated immunity against  
 PT target cells, such as tumor cells that express C35 gene -

XX  
 XX Disclosure; Page 61; 331pp; English.

XX  
 XX The present invention relates to human C35 (see AAG78997). C35 is a novel  
 CC tumour antigen that is overexpressed in human breast and bladder  
 CC carcinoma. C35 is thought to be a promising candidate for tumour  
 CC immunotherapy, in immunogenic compositions and vaccines, to induce  
 CC antibody and cell-mediated immunity against target cells such as tumour  
 CC cells that express C35 genes. The present sequence is a C35 peptide  
 CC epitope. This peptide is predicted to be a major histocompatibility  
 CC complex (MHC) binding peptide.

XX  
 XX Sequence 15 AA;

Query Match 11.5%; Score 15; DB 22; Length 15;

Best Local Similarity 100.0%; Pred. No. 1.5e-07;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 36 GSGVRIVECEPCG 50.  
 |||||  
 1 GSGVRIVECEPCG 15

RESULT 28

ABB12837  
 ID ABB12837 standard; Peptide; 15 AA.

XX  
 XX ABB12837;

XX  
 XX 22-JAN-2002 (first entry)

XX  
 XX Human C35 peptide epitope #341.

KM Human; C35; cytostatic; gene therapy; vaccine; tumour antigen;  
 KM breast cancer; bladder cancer; tumour immunotherapy; epitope;  
 KM major histocompatibility complex binding peptide; MHC.

XX  
 XX Homo sapiens.

XX  
 XX WO200174859-A2.

XX  
 XX 11-OCT-2001.

PF 04-APR-2001; 2001WO-US10855.  
 XX  
 PR 04-APR-2000; 2000US-194463P.  
 XX  
 PA (UYRP ) UNIV ROCHESTER.  
 XX  
 PI Zauderer M, Evans EE, Borrello MA;  
 XX  
 DR WPI; 2001-626383/72.  
 XX  
 PT Novel C35 polypeptides and C35 genes useful in immunogenic compositions  
 PT and vaccines, for inducing antibody and cell-mediated immunity against  
 PT target cells, such as tumor cells that express C35 gene -  
 XX  
 PS Disclosure; Page 61; 331pp; English.  
 XX  
 CC The present invention relates to human C35 (see AAG78997). C35 is a novel  
 CC tumour antigen that is overexpressed in human breast and bladder  
 CC carcinoma. C35 is thought to be a promising candidate for tumour  
 CC immunotherapy, in immunogenic compositions and vaccines, to induce  
 CC antibody and cell-mediated immunity against target cells such as tumour  
 CC cells that express C35 genes. The present sequence is a C35 peptide  
 CC epitope. This peptide is predicted to be a major histocompatibility  
 CC complex (MHC) binding peptide.  
 CC  
 SQ Sequence 15 AA;  
 XX  
 Query Match 11.5%; Score 15; DB 22; Length 15;  
 Best Local Similarity 100.0%; Pred. No. 1.5e-07;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 78 TGAFEIRINGOLVRS 92  
 DB 1 TGAFEIRINGOLVRS 15  
 XX  
 RESULT 29  
 ABB12838  
 ID ABB12838 standard; Peptide; 15 AA.  
 XX  
 AC ABB12838;  
 XX  
 DT 22-JAN-2002 (first entry)  
 XX  
 DE Human C35 peptide epitope #342.  
 XX  
 KW Human; C35; cytostatic; gene therapy; vaccine; tumour antigen;  
 KW breast cancer; bladder cancer; tumour immunotherapy; epitope;  
 KW major histocompatibility complex binding peptide; MHC.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200174859-A2.  
 XX  
 PD 11-OCT-2001.  
 XX  
 PF 04-APR-2001; 2001WO-US10855.  
 XX  
 PR 04-APR-2000; 2000US-194463P.  
 XX  
 PA (UYRP ) UNIV ROCHESTER.  
 XX  
 PI Zauderer M, Evans EE, Borrello MA;  
 XX  
 DR WPI; 2001-626383/72.  
 XX  
 PT Novel C35 polypeptides and C35 genes useful in immunogenic compositions  
 PT and vaccines, for inducing antibody and cell-mediated immunity against  
 PT target cells, such as tumor cells that express C35 gene -  
 XX  
 PS Disclosure; Page 61; 331pp; English.  
 XX  
 CC The present invention relates to human C35 (see AAG78997). C35 is a novel  
 CC tumour antigen that is overexpressed in human breast and bladder  
 CC carcinoma. C35 is thought to be a promising candidate for tumour  
 CC immunotherapy, in immunogenic compositions and vaccines, to induce  
 CC antibody and cell-mediated immunity against target cells such as tumour  
 CC cells that express C35 genes. The present sequence is a C35 peptide  
 CC epitope. This peptide is predicted to be a major histocompatibility  
 CC complex (MHC) binding peptide.  
 CC  
 SQ Sequence 15 AA;  
 XX  
 Query Match 11.5%; Score 15; DB 22; Length 15;  
 Best Local Similarity 100.0%; Pred. No. 1.5e-07;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 63 KEQYPGIEISRLGG 77  
 DB 1 KEQYPGIEISRLGG 15

CC tumour antigen that is overexpressed in human breast and bladder  
 CC carcinoma. C35 is thought to be a promising candidate for tumour  
 CC immunotherapy, in immunogenic compositions and vaccines, to induce  
 CC antibody and cell-mediated immunity against target cells such as tumour  
 CC cells that express C35 genes. The present sequence is a C35 peptide  
 CC epitope. This peptide is predicted to be a major histocompatibility  
 CC complex (MHC) binding peptide.  
 CC  
 SQ Sequence 15 AA;  
 XX  
 Query Match 11.5%; Score 15; DB 22; Length 15;  
 Best Local Similarity 100.0%; Pred. No. 1.5e-07;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 48 PCGFATYELIASAV 62  
 DB 1 PCGFATYELIASAV 15  
 XX  
 RESULT 30  
 ABB12839  
 ID ABB12839 standard; Peptide; 15 AA.  
 XX  
 AC ABB12839;  
 XX  
 DT 22-JAN-2002 (first entry)  
 XX  
 DE Human C35 peptide epitope #343.  
 XX  
 KW Human; C35; cytostatic; gene therapy; vaccine; tumour antigen;  
 KW breast cancer; bladder cancer; tumour immunotherapy; epitope;  
 KW major histocompatibility complex binding peptide; MHC.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200174859-A2.  
 XX  
 PD 11-OCT-2001.  
 XX  
 PF 04-APR-2001; 2001WO-US10855.  
 XX  
 PR 04-APR-2000; 2000US-194463P.  
 XX  
 PA (UYRP ) UNIV ROCHESTER.  
 XX  
 PI Zauderer M, Evans EE, Borrello MA;  
 XX  
 DR WPI; 2001-626383/72.  
 XX  
 PT Novel C35 polypeptides and C35 genes useful in immunogenic compositions  
 PT and vaccines, for inducing antibody and cell-mediated immunity against  
 PT target cells, such as tumor cells that express C35 gene -  
 XX  
 PS Disclosure; Page 61; 331pp; English.  
 XX  
 CC The present invention relates to human C35 (see AAG78997). C35 is a novel  
 CC tumour antigen that is overexpressed in human breast and bladder  
 CC carcinoma. C35 is thought to be a promising candidate for tumour  
 CC immunotherapy, in immunogenic compositions and vaccines, to induce  
 CC antibody and cell-mediated immunity against target cells such as tumour  
 CC cells that express C35 genes. The present sequence is a C35 peptide  
 CC epitope. This peptide is predicted to be a major histocompatibility  
 CC complex (MHC) binding peptide.  
 CC  
 SQ Sequence 15 AA;  
 XX  
 Query Match 11.5%; Score 15; DB 22; Length 15;  
 Best Local Similarity 100.0%; Pred. No. 1.5e-07;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 63 KEQYPGIEISRLGG 77  
 DB 1 KEQYPGIEISRLGG 15

	RESULT	31	
XX	ABBI12840	ID	ABBI12840 standard; Peptide; 15 AA.
AC	ABBI12840;	XX	
XX		DT	22-JAN-2002 (first entry)
XX		DE	Human C35 peptide epitope #344.
XX		KW	Human; C35; cytostatic; gene therapy; vaccine; tumour antigen; breast cancer; bladder cancer; tumour immunotherapy; epitope; major histocompatibility complex binding peptide; MHC.
XV	Homo sapiens.	OS	
XX	WO200174859-A2.	PB	
PD	11-OCT-2001.	PF	
PR	04-APR-2001; 2001MO-US10855.	PP	
XX	04-APR-2000; 2000US--194463P.	PT	(UVRP ) UNIV ROCHESTER.
PA	Zauderer M, Evans EE, Borrello MA,	PI	
XX	WPt; 2001-626383/72.	DR	
PS	Novel C35 polypeptides and C35 genes useful in immunogenic compositions and vaccines, for inducing antibody and cell-mediated immunity against target cells, such as tumor cells that express C35 gene -  Disclosure; Page 61; 331pp; English.	FT	
XX	The present invention relates to human C35 (see AAG78997). C35 is a novel tumour antigen that is overexpressed in human breast and bladder carcinoma. C35 is thought to be a promising candidate for tumour immunotherapy, in immunogenic compositions and vaccines, to induce antibody and cell-mediated immunity against target cells such as tumour cells that express C35 genes. The present sequence is a C35 peptide epitope. This peptide is predicted to be a major histocompatibility complex (MHC) binding peptide.  Sequence    15 AA;	CC	
SQ	Query Match	SC	
	Best Local Similarity     11.5%; Score 15; DB 22; Length 15;		
	Matches    15; Conservative     0; Pred. No. 1.5e-07; Indels       0; Gaps      0;		
OY	80 AFEEINGQLVFESKL    94		
DB	1 AFEEINGQLVFESKL    15		
	RESULT    32		
	ABBI12841		
ID	ABBI12841 standard; Peptide; 15 AA.		
XX			
AC	ABBI12841;		
XX			
DT	22-JAN-2002 (first entry)		
XX			
DE	Human C35 peptide epitope #345.		
XX			
KW	Human; C35; cytostatic; gene therapy; vaccine; tumour antigen; breast cancer; bladder cancer; tumour immunotherapy; epitope; major histocompatibility complex binding peptide; MHC.		
XV	Homo sapiens.		
OS			

[illegible]

```

PT target cells, such as tumor cells that express C35 gene
XX
XX Disclosure; Page 61; 331pp; English.
PS
XX The present invention relates to human C35 (see AAG78997). C35 is a novel
CC tumour antigen that is overexpressed in human breast and bladder
CC carcinoma. C35 is thought to be a promising candidate for tumour
CC immunotherapy, in immunogenic compositions and vaccines, to induce
CC antibody and cell-mediated immunity against target cells such as tumour
CC cells that express C35 genes. The present sequence is a C35 peptide
CC epitope. This peptide is predicted to be a major histocompatibility
CC complex (MHC) binding peptide.
XX
XX Sequence 15 AA;
SQ
Query Match 11.5%; Score 15; DB 22; Length 15;
Best Local Similarity 100.0%; Pred. No. 1.5e-07;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 50 GPEATYLELASAYKE 64
DB 1 GPEATYLELASAYKE 15
RESULT 34
ABBI2843
ID ABBI2843 standard; Peptide; 15 AA.
XX
XX ABBI2843;
AC
XX 22-JAN-2002 (first entry)
DT
XX
XX Human C35 peptide epitope #347.
DE
XX
XX Human; C35; cytostatic; gene therapy; vaccine; tumour antigen;
KM breast cancer; bladder cancer; tumour immunotherapy; epitope;
KW major histocompatibility complex binding peptide; MHC.
XX
XX Homo sapiens.
OS
XX
XX WO200174859-A2.
PN
XX
XX 11-OCT-2001.
PD
XX
XX 04-APR-2001; 2001WO-US10855.
PF
XX
XX 04-APR-2000; 2000US-194463P.
PR
XX
XX (UYRP ) UNIV ROCHESTER.
PA
XX
XX Zauderer M, Evans EE, Borrello MA;
PI
XX
XX WPI; 2001-626383/72.
DR
XX
XX Novel C35 polypeptides and C35 genes useful in immunogenic compositions
PT and vaccines, for inducing antibody and cell-mediated immunity against
PT target cells, such as tumor cells that express C35 gene
PS
XX Disclosure; Page 61; 331pp; English.
PS
XX The present invention relates to human C35 (see AAG78997). C35 is a novel
CC tumour antigen that is overexpressed in human breast and bladder
CC carcinoma. C35 is thought to be a promising candidate for tumour
CC immunotherapy, in immunogenic compositions and vaccines, to induce
CC antibody and cell-mediated immunity against target cells such as tumour
CC cells that express C35 genes. The present sequence is a C35 peptide
CC epitope. This peptide is predicted to be a major histocompatibility
CC complex (MHC) binding peptide.
XX
XX Sequence 15 AA;
SQ
Query Match 11.5%; Score 15; DB 22; Length 15;
Best Local Similarity 100.0%; Pred. No. 1.5e-07;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 70 IEISRLGGTGAFEIE 84
DB 1 IEISRLGGTGAFEIE 15
RESULT 35
ABBI2844
ID ABBI2844 standard; Peptide; 15 AA.
XX
XX ABBI2844;
AC
XX
XX 22-JAN-2002 (first entry)
DT
XX
XX Human C35 peptide epitope #348.
DE
XX
XX Human; C35; cytostatic; gene therapy; vaccine; tumour antigen;
KM breast cancer; bladder cancer; tumour immunotherapy; epitope;
KW major histocompatibility complex binding peptide; MHC.
XX
XX Homo sapiens.
OS
XX
XX WO200174859-A2.
PN
XX
XX 11-OCT-2001.
PD
XX
XX 04-APR-2001; 2001WO-US10855.
PF
XX
XX 04-APR-2000; 2000US-194463P.
PR
XX
XX (UYRP ) UNIV ROCHESTER.
PA
XX
XX Zauderer M, Evans EE, Borrello MA;
PI
XX
XX WPI; 2001-626383/72.
DR
XX
XX Novel C35 polypeptides and C35 genes useful in immunogenic compositions
PT and vaccines, for inducing antibody and cell-mediated immunity against
PT target cells, such as tumor cells that express C35 gene
PS
XX Disclosure; Page 61; 331pp; English.
PS
XX The present invention relates to human C35 (see AAG78997). C35 is a novel
CC tumour antigen that is overexpressed in human breast and bladder
CC carcinoma. C35 is thought to be a promising candidate for tumour
CC immunotherapy, in immunogenic compositions and vaccines, to induce
CC antibody and cell-mediated immunity against target cells such as tumour
CC cells that express C35 genes. The present sequence is a C35 peptide
CC epitope. This peptide is predicted to be a major histocompatibility
CC complex (MHC) binding peptide.
XX
XX Sequence 15 AA;
SQ
Query Match 11.5%; Score 15; DB 22; Length 15;
Best Local Similarity 100.0%; Pred. No. 1.5e-07;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 106 IEAIRRASNGETLEK 120
DB 1 IEAIRRASNGETLEK 15
RESULT 36
ABBI2845
ID ABBI2845 standard; Peptide; 15 AA.
XX
XX ABBI2845;
AC
XX
XX 22-JAN-2002 (first entry)
DT
XX
XX Human C35 peptide epitope #349.
DE
XX

```

KM Human; C35; cytostatic; gene therapy; vaccine; tumour antigen;  
 KW breast cancer; bladder cancer; tumour immunotherapy; epitope;  
 KW major histocompatibility complex binding peptide; MHC.  
 XX Homo sapiens.  
 XX WO200174859-A2.  
 PN  
 XX  
 XX 11-OCT-2001.  
 PD  
 XX  
 XX 04-APR-2001; 2001WO-US10855.  
 PF  
 XX  
 XX 04-APR-2000; 2000US-194463P.  
 PR  
 XX  
 XX (UYRP ) UNIV ROCHESTER.  
 PA  
 XX  
 XX Zauderer M, Evans EE, Borrello MA;  
 PI  
 XX WPI, 2001-626383/72.  
 DR  
 XX  
 XX Novel C35 polypeptides and C35 genes useful in immunogenic compositions  
 PT and vaccines, for inducing antibody and cell-mediated immunity against  
 PT target cells, such as tumor cells that express C35 gene -  
 PT  
 XX  
 XX Disclosure; Page 61; 331pp; English.  
 PS  
 XX  
 XX The present invention relates to human C35 (see AAG78997). C35 is a novel  
 CC tumour antigen that is overexpressed in human breast and bladder  
 CC carcinoma. C35 is thought to be a promising candidate for tumour  
 CC immunotherapy, in immunogenic compositions and vaccines, to induce  
 CC antibody and cell-mediated immunity against target cells such as tumour  
 CC cells that express C35 genes. The present sequence is a C35 peptide  
 CC epitope. This peptide is predicted to be a major histocompatibility  
 CC complex (MHC) binding peptide.  
 CC  
 XX  
 XX Sequence 15 AA;  
 SQ  
 Query Match 11.5%; Score 15; DB 22; Length 15;  
 Best Local Similarity 100.0%; Pred. No. 1.5e-07;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 115 GRTLEKITSRPPCV 129  
 DB 1 GRTLEKITSRPPCV 15  
 RESULT 37  
 ABB12846  
 ID ABB12846 standard; Peptide; 15 AA.  
 XX  
 XX ABB12846;  
 AC  
 XX  
 XX 22-JAN-2002 (first entry)  
 DT  
 XX  
 XX Human C35 peptide epitope #350.  
 DE  
 XX  
 XX Human; C35; cytostatic; gene therapy; vaccine; tumour antigen;  
 KW breast cancer; bladder cancer; tumour immunotherapy; epitope;  
 KW major histocompatibility complex binding peptide; MHC.  
 XX  
 XX Homo sapiens.  
 OS  
 XX  
 XX WO200174859-A2.  
 PN  
 XX  
 XX 11-OCT-2001.  
 PD  
 XX  
 XX 04-APR-2001; 2001WO-US10855.  
 PF  
 XX  
 XX 04-APR-2000; 2000US-194463P.  
 PR  
 XX  
 XX (UYRP ) UNIV ROCHESTER.  
 PA  
 XX  
 XX Zauderer M, Evans EE, Borrello MA;  
 PI

XX  
 DR WPI, 2001-626383/72.  
 XX  
 XX Novel C35 polypeptides and C35 genes useful in immunogenic compositions  
 PT and vaccines, for inducing antibody and cell-mediated immunity against  
 PT target cells, such as tumor cells that express C35 gene -  
 PT  
 XX  
 XX Disclosure; Page 61; 331pp; English.  
 PS  
 XX  
 XX The present invention relates to human C35 (see AAG78997). C35 is a novel  
 CC tumour antigen that is overexpressed in human breast and bladder  
 CC carcinoma. C35 is thought to be a promising candidate for tumour  
 CC immunotherapy, in immunogenic compositions and vaccines, to induce  
 CC antibody and cell-mediated immunity against target cells such as tumour  
 CC cells that express C35 genes. The present sequence is a C35 peptide  
 CC epitope. This peptide is predicted to be a major histocompatibility  
 CC complex (MHC) binding peptide.  
 CC  
 XX  
 XX Sequence 15 AA;  
 SQ  
 Query Match 11.5%; Score 15; DB 22; Length 15;  
 Best Local Similarity 100.0%; Pred. No. 1.5e-07;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 47 EPCGEATYLELASA 61  
 DB 1 EPCGEATYLELASA 15  
 RESULT 38  
 ABB12847  
 ID ABB12847 standard; Peptide; 15 AA.  
 XX  
 XX ABB12847;  
 AC  
 XX  
 XX 22-JAN-2002 (first entry)  
 DT  
 XX  
 XX Human C35 peptide epitope #351.  
 DE  
 XX  
 XX Human; C35; cytostatic; gene therapy; vaccine; tumour antigen;  
 KW breast cancer; bladder cancer; tumour immunotherapy; epitope;  
 KW major histocompatibility complex binding peptide; MHC.  
 XX  
 XX Homo sapiens.  
 OS  
 XX  
 XX WO200174859-A2.  
 PN  
 XX  
 XX 11-OCT-2001.  
 PD  
 XX  
 XX 04-APR-2001; 2001WO-US10855.  
 PF  
 XX  
 XX 04-APR-2000; 2000US-194463P.  
 PR  
 XX  
 XX (UYRP ) UNIV ROCHESTER.  
 PA  
 XX  
 XX Zauderer M, Evans EE, Borrello MA;  
 PI  
 XX WPI, 2001-626383/72.  
 DR  
 XX  
 XX Novel C35 polypeptides and C35 genes useful in immunogenic compositions  
 PT and vaccines, for inducing antibody and cell-mediated immunity against  
 PT target cells, such as tumor cells that express C35 gene -  
 PT  
 XX  
 XX Disclosure; Page 61; 331pp; English.  
 PS  
 XX  
 XX The present invention relates to human C35 (see AAG78997). C35 is a novel  
 CC tumour antigen that is overexpressed in human breast and bladder  
 CC carcinoma. C35 is thought to be a promising candidate for tumour  
 CC immunotherapy, in immunogenic compositions and vaccines, to induce  
 CC antibody and cell-mediated immunity against target cells such as tumour  
 CC cells that express C35 genes. The present sequence is a C35 peptide  
 CC epitope. This peptide is predicted to be a major histocompatibility  
 CC complex (MHC) binding peptide.

```

XX Sequence 15 AA;
SQ
Query Match 11.5%; Score 15; DB 22; Length 15;
Best Local Similarity 100.0%; Pred. No. 1.5e-07;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 65 QYPGIEIESRLGQTG 79
Db 1 QYPGIEIESRLGQTG 15

RESULT 39
ABBI2848
ID ABBI2848 standard; Peptide; 15 AA.
XX
AC ABBI2848;
XX
DT 22-JAN-2002 (first entry)
XX
DE Human C35 peptide epitope #352.
XX
KW Human; C35; cytostatic; gene therapy; vaccine; tumour antigen;
KW breast cancer; bladder cancer; tumour immunotherapy; epitope;
KW major histocompatibility complex binding peptide; MHC.
XX
OS Homo sapiens.
XX
PN WO200174859-A2.
XX
PD 11-OCT-2001.
XX
PF 04-APR-2001; 2001WO-US10855.
XX
PR 04-APR-2000; 2000US-194463P.
XX
PA (UYRP ) UNIV ROCHESTER.
XX
PI Zauderer M, Evans EE, Borrello MA;
XX
PI WPI; 2001-626383/72.
XX
DR Novel C35 polypeptides and C35 genes useful in immunogenic compositions
XX and vaccines, for inducing antibody and cell-mediated immunity against
XX target cells, such as tumor cells that express C35 gene
XX
PS Disclosure; Page 61; 331pp; English.
XX
CC The present invention relates to human C35 (see AAG78997). C35 is a novel
XX tumour antigen that is overexpressed in human breast and bladder
XX carcinoma. C35 is thought to be a promising candidate for tumour
XX immunotherapy, in immunogenic compositions and vaccines, to induce
XX antibody and cell-mediated immunity against target cells such as tumour
XX cells that express C35 genes. The present sequence is a C35 peptide
XX epitope. This peptide is predicted to be a major histocompatibility
XX complex (MHC) binding peptide.
XX
SQ Sequence 15 AA;
QY
Query Match 11.5%; Score 15; DB 22; Length 15;
Best Local Similarity 100.0%; Pred. No. 1.5e-07;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 74 RLGGTGAFEIRINQ 88
Db 1 RLGGTGAFEIRINQ 15

RESULT 40
ABBI2849
ID ABBI2849 standard; Peptide; 15 AA.
XX
AC ABBI2849;
XX

```

```

XX
DT 22-JAN-2002 (first entry)
XX
DE Human C35 peptide epitope #353.
XX
KW Human; C35; cytostatic; gene therapy; vaccine; tumour antigen;
KW breast cancer; bladder cancer; tumour immunotherapy; epitope;
KW major histocompatibility complex binding peptide; MHC.
XX
OS Homo sapiens.
XX
PN WO200174859-A2.
XX
PD 11-OCT-2001.
XX
PF 04-APR-2001; 2001WO-US10855.
XX
PR 04-APR-2000; 2000US-194463P.
XX
PA (UYRP ) UNIV ROCHESTER.
XX
PI Zauderer M, Evans EE, Borrello MA;
XX
PI WPI; 2001-626383/72.
XX
DR Novel C35 polypeptides and C35 genes useful in immunogenic compositions
XX and vaccines, for inducing antibody and cell-mediated immunity against
XX target cells, such as tumor cells that express C35 gene
XX
PS Disclosure; Page 62; 331pp; English.
XX
CC The present invention relates to human C35 (see AAG78997). C35 is a novel
XX tumour antigen that is overexpressed in human breast and bladder
XX carcinoma. C35 is thought to be a promising candidate for tumour
XX immunotherapy, in immunogenic compositions and vaccines, to induce
XX antibody and cell-mediated immunity against target cells such as tumour
XX cells that express C35 genes. The present sequence is a C35 peptide
XX epitope. This peptide is predicted to be a major histocompatibility
XX complex (MHC) binding peptide.
XX
SQ Sequence 15 AA;
QY
Query Match 11.5%; Score 15; DB 22; Length 15;
Best Local Similarity 100.0%; Pred. No. 1.5e-07;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 82 EIRINGQLVFSKLEN 96
Db 1 EIRINGQLVFSKLEN 15

RESULT 41
ABBI2850
ID ABBI2850 standard; Peptide; 15 AA.
XX
AC ABBI2850;
XX
DT 22-JAN-2002 (first entry)
XX
DE Human C35 peptide epitope #354.
XX
KW Human; C35; cytostatic; gene therapy; vaccine; tumour antigen;
KW breast cancer; bladder cancer; tumour immunotherapy; epitope;
KW major histocompatibility complex binding peptide; MHC.
XX
OS Homo sapiens.
XX
PN WO200174859-A2.
XX
PD 11-OCT-2001.
XX
PF 04-APR-2001; 2001WO-US10855.
XX

```

PR 04-APR-2000; 2000US-194463P.  
XX (UYRP ) UNIV ROCHESTER.  
XX Zauderer M, Evans EE, Borrello MA;  
XX WPI; 2001-626383/72.  
XX  
XX Novel C35 polypeptides and C35 genes useful in immunogenic compositions  
XX and vaccines, for inducing antibody and cell-mediated immunity against  
XX target cells, such as tumor cells that express C35 gene -  
XX  
XX  
XX Disclosure; Page 62; 331pp; English.  
XX  
XX The present invention relates to human C35 (see AAG78997). C35 is a novel  
XX tumour antigen that is overexpressed in human breast and bladder  
XX carcinoma. C35 is thought to be a promising candidate for tumour  
XX immunotherapy, in immunogenic compositions and vaccines, to induce  
XX antibody and cell-mediated immunity against target cells such as tumour  
XX cells that express C35 genes. The present sequence is a C35 peptide  
XX epitope. This peptide is predicted to be a major histocompatibility  
XX complex (MHC) binding peptide.  
XX  
XX  
XX Sequence 15 AA:  
SQ  
Query Match 11.5%; Score 15; DB 22; Length 15;  
Best Local Similarity 100.0%; Pred. No. 1.5e-07;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 83 IIRINGQLVPSKLENG 97  
Db 1 IIRINGQLVPSKLENG 15  
RESULT 42  
ABBI2851  
ID ABBI2851 standard; Peptide; 15 AA.  
XX  
XX ABB12851;  
XX  
XX 22-JAN-2002 (first entry)  
XX  
XX Human C35 peptide epitope #355.  
XX  
XX Human; C35; cytostatic; gene therapy; vaccine; tumour antigen;  
XX breast cancer; bladder cancer; tumour immunotherapy; epitope;  
XX major histocompatibility complex binding peptide; MHC.  
XX  
XX Homo sapiens.  
XX  
XX WO200174859-A2.  
XX  
XX 11-OCT-2001.  
XX  
XX 04-APR-2001; 2001WO-US10855.  
XX  
XX 04-APR-2000; 2000US-194463P.  
XX  
XX (UYRP ) UNIV ROCHESTER.  
XX  
XX Zauderer M, Evans EE, Borrello MA;  
XX WPI; 2001-626383/72.  
XX  
XX Novel C35 polypeptides and C35 genes useful in immunogenic compositions  
XX and vaccines, for inducing antibody and cell-mediated immunity against  
XX target cells, such as tumor cells that express C35 gene -  
XX  
XX Disclosure; Page 62; 331pp; English.  
XX  
XX The present invention relates to human C35 (see AAG78997). C35 is a novel  
XX tumour antigen that is overexpressed in human breast and bladder  
XX carcinoma. C35 is thought to be a promising candidate for tumour

CC immunotherapy, in immunogenic compositions and vaccines, to induce  
CC antibody and cell-mediated immunity against target cells such as tumour  
CC cells that express C35 genes. The present sequence is a C35 peptide  
CC epitope. This peptide is predicted to be a major histocompatibility  
CC complex (MHC) binding peptide.  
XX  
XX  
XX Sequence 15 AA;  
SQ  
Query Match 11.5%; Score 15; DB 22; Length 15;  
Best Local Similarity 100.0%; Pred. No. 1.5e-07;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 84 IIRINGQLVPSKLENG 98  
Db 1 IIRINGQLVPSKLENG 15  
RESULT 43  
ABBI2852  
ID ABBI2852 standard; Peptide; 15 AA.  
XX  
XX ABB12852;  
XX  
XX 22-JAN-2002 (first entry)  
XX  
XX Human C35 peptide epitope #356.  
XX  
XX Human; C35; cytostatic; gene therapy; vaccine; tumour antigen;  
XX breast cancer; bladder cancer; tumour immunotherapy; epitope;  
XX major histocompatibility complex binding peptide; MHC.  
XX  
XX Homo sapiens.  
XX  
XX WO200174859-A2.  
XX  
XX 11-OCT-2001.  
XX  
XX 04-APR-2001; 2001WO-US10855.  
XX  
XX 04-APR-2000; 2000US-194463P.  
XX  
XX (UYRP ) UNIV ROCHESTER.  
XX  
XX Zauderer M, Evans EE, Borrello MA;  
XX WPI; 2001-626383/72.  
XX  
XX Novel C35 polypeptides and C35 genes useful in immunogenic compositions  
XX and vaccines, for inducing antibody and cell-mediated immunity against  
XX target cells, such as tumor cells that express C35 gene -  
XX  
XX Disclosure; Page 62; 331pp; English.  
XX  
XX The present invention relates to human C35 (see AAG78997). C35 is a novel  
XX tumour antigen that is overexpressed in human breast and bladder  
XX carcinoma. C35 is thought to be a promising candidate for tumour  
XX immunotherapy, in immunogenic compositions and vaccines, to induce  
XX antibody and cell-mediated immunity against target cells such as tumour  
XX cells that express C35 genes. The present sequence is a C35 peptide  
XX epitope. This peptide is predicted to be a major histocompatibility  
XX complex (MHC) binding peptide.  
XX  
XX  
XX Sequence 15 AA;  
SQ  
Query Match 11.5%; Score 15; DB 22; Length 15;  
Best Local Similarity 100.0%; Pred. No. 1.5e-07;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 100 PYEKDLIETATRRASN 114  
Db 1 PYEKDLIETATRRASN 15

```

XX 11-OCT-2001.
PD
XX
XX 04-APR-2001; 2001WO-US10855.
PF
XX 04-APR-2000; 2000US-194463P.
PR
XX (UYRP ) UNIV ROCHESTER.
PA
XX Zauderer M, Evans EE, Borrello MA;
PI
DR WPI; 2001-626383/72.
XX
XX Novel C35 polypeptides and C35 genes useful in immunogenic compositions
PT and vaccines, for inducing antibody and cell-mediated immunity against
PT target cells, such as tumor cells that express C35 gene
XX
XX Disclosure; Page 62; 31pp; English.
PS
XX The present invention relates to human C35 (see AAG78997). C35 is a novel
CC tumour antigen that is overexpressed in human breast and bladder
CC carcinoma. C35 is thought to be a promising candidate for tumour
CC immunotherapy, in immunogenic compositions and vaccines, to induce
CC antibody and cell-mediated immunity against target cells such as tumour
CC cells that express C35 genes. The present sequence is a C35 peptide
CC epitope. This peptide is predicted to be a major histocompatibility
CC complex (MHC) binding peptide.
SQ
XX Sequence 15 AA;
XX
XX Query Match 11.5%; Score 15; DB 22; Length 15;
XX Best Local Similarity 100.0%; Pred. No. 1.5e-07;
XX Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 51 FEATYLELASAVKEQ 65
XX ||||||||||||
DB 1 FEATYLELASAVKEQ 15
XX
XX RESULT 46
XX ABB12855
XX ID ABB12855 standard; Peptide; 15 AA.
XX
XX ABB12855;
XX AC
XX DT 22-JAN-2002 (first entry)
XX
XX Human C35 peptide epitope #359.
XX
XX Human; C35; cytostatic; gene therapy; vaccine; tumour antigen;
XX breast cancer; bladder cancer; tumour immunotherapy; epitope;
XX major histocompatibility complex binding peptide; MHC.
XX
XX Homo sapiens.
XX
XX WO200174859-A2.
XX PN
XX 11-OCT-2001.
XX
XX 04-APR-2001; 2001WO-US10855.
XX PF
XX 04-APR-2000; 2000US-194463P.
XX PR
XX (UYRP ) UNIV ROCHESTER.
XX PA
XX Zauderer M, Evans EE, Borrello MA;
XX PI
XX WPI; 2001-626383/72.
XX
XX Novel C35 polypeptides and C35 genes useful in immunogenic compositions
PT and vaccines, for inducing antibody and cell-mediated immunity against
PT target cells, such as tumor cells that express C35 gene
XX

```



PS Disclosure; Page 62; 331pp; English.

XX The present invention relates to human C35 (see AAG78997). C35 is a novel  
 CC tumour antigen that is overexpressed in human breast and bladder  
 CC carcinoma. C35 is thought to be a promising candidate for tumour  
 CC immunotherapy, in immunogenic compositions and vaccines, to induce  
 CC antibody and cell-mediated immunity against target cells such as tumour  
 CC cells that express C35 genes. The present sequence is a C35 peptide  
 CC epitope. This peptide is predicted to be a major histocompatibility  
 CC complex (MHC) binding peptide.

XX  
 SQ Sequence 15 AA;

Query Match 11.5%; Score 15; DB 22; Length 15;  
 Best Local Similarity 100.0%; Pred. No. 1.5e-07;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 90 VFSKLENGGFPYKED 104  
 Db 1 VFSKLENGGFPYKED 15

## RESULT 47

ABBI2856  
 ID ABBI2856 standard; Peptide; 15 AA.

AC ABBI2856;

DT 22-JAN-2002 (first entry)

DE Human C35 peptide epitope #360.

XX Human; C35; cytostatic; gene therapy; vaccine; tumour antigen;

KW breast cancer; bladder cancer; tumour immunotherapy; epitope;

KM major histocompatibility complex binding peptide; MHC.

XX Homo sapiens.

PN WO200174859-A2.

PD 11-OCT-2001.

PF 04-APR-2001; 2001WO-US10855.

PR 04-APR-2000; 2000US-194463P.

PA (UYRP ) UNIV ROCHESTER.

PI Zauderer M, Evans EE, Borrello MA;

DR WPI; 2001-626383/72.

XX Novel C35 polypeptides and C35 genes useful in immunogenic compositions  
 PT and vaccines, for inducing antibody and cell-mediated immunity against  
 PT target cells, such as tumor cells that express C35 gene -

PS Disclosure; Page 62; 331pp; English.

XX The present invention relates to human C35 (see AAG78997). C35 is a novel  
 CC tumour antigen that is overexpressed in human breast and bladder  
 CC carcinoma. C35 is thought to be a promising candidate for tumour  
 CC immunotherapy, in immunogenic compositions and vaccines, to induce  
 CC antibody and cell-mediated immunity against target cells such as tumour  
 CC cells that express C35 genes. The present sequence is a C35 peptide  
 CC epitope. This peptide is predicted to be a major histocompatibility  
 CC complex (MHC) binding peptide.

XX Sequence 15 AA;

Query Match 11.5%; Score 15; DB 22; Length 15;  
 Best Local Similarity 100.0%; Pred. No. 1.5e-07;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 103 KDLIEAIRRANGSET 117  
 Db 1 KDLIEAIRRANGSET 15

## RESULT 48

ABBI2857  
 ID ABBI2857 standard; Peptide; 15 AA.

AC ABBI2857;

DT 22-JAN-2002 (first entry)

DE Human C35 peptide epitope #361.

XX Human; C35; cytostatic; gene therapy; vaccine; tumour antigen;

KW breast cancer; bladder cancer; tumour immunotherapy; epitope;

KM major histocompatibility complex binding peptide; MHC.

XX Homo sapiens.

PN WO200174859-A2.

PD 11-OCT-2001.

PF 04-APR-2001; 2001WO-US10855.

PR 04-APR-2000; 2000US-194463P.

PA (UYRP ) UNIV ROCHESTER.

PI Zauderer M, Evans EE, Borrello MA;

DR WPI; 2001-626383/72.

XX Novel C35 polypeptides and C35 genes useful in immunogenic compositions  
 PT and vaccines, for inducing antibody and cell-mediated immunity against  
 PT target cells, such as tumor cells that express C35 gene -

PS Disclosure; Page 62; 331pp; English.

XX The present invention relates to human C35 (see AAG78997). C35 is a novel  
 CC tumour antigen that is overexpressed in human breast and bladder  
 CC carcinoma. C35 is thought to be a promising candidate for tumour  
 CC immunotherapy, in immunogenic compositions and vaccines, to induce  
 CC antibody and cell-mediated immunity against target cells such as tumour  
 CC cells that express C35 genes. The present sequence is a C35 peptide  
 CC epitope. This peptide is predicted to be a major histocompatibility  
 CC complex (MHC) binding peptide.

XX Sequence 15 AA;

Query Match 11.5%; Score 15; DB 22; Length 15;  
 Best Local Similarity 100.0%; Pred. No. 1.5e-07;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 107 EAIRRANGSETLEKI 121  
 Db 1 EAIRRANGSETLEKI 15

## RESULT 49

ABBI2858  
 ID ABBI2858 standard; Peptide; 15 AA.

AC ABBI2858;

DT 22-JAN-2002 (first entry)

DE Human C35 peptide epitope #362.

XX Human; C35; cytostatic; gene therapy; vaccine; tumour antigen;

KW breast cancer; bladder cancer; tumour immunotherapy; epitope;

KW major histocompatibility complex binding peptide; MHC.  
XX  
OS Homo sapiens.  
XX  
PN WO200174859-A2.  
XX  
PD 11-OCT-2001.  
XX  
PF 04-APR-2001; 2001WO-US10855.  
XX  
PR 04-APR-2000; 2000US-194463P.  
XX  
PA (UYRP ) UNIV ROCHESTER.  
XX  
PI Zauderer M, Evans EE, Borrello MA;  
XX WPI; 2001-626383/72.  
DR  
XX  
XX Novel C35 polypeptides and C35 genes useful in immunogenic compositions  
PT and vaccines, for inducing antibody and cell-mediated immunity against  
PT target cells, such as tumor cells that express C35 gene -  
XX  
PS Disclosure; Page 62; 33pp; English.  
XX  
XX The present invention relates to human C35 (see AAG78997). C35 is a novel  
CC tumour antigen that is overexpressed in human breast and bladder  
CC carcinoma. C35 is thought to be a promising candidate for tumour  
CC immunotherapy, in immunogenic compositions and vaccines, to induce  
CC antibody and cell-mediated immunity against target cells such as tumour  
CC cells that express C35 genes. The present sequence is a C35 peptide  
CC epitope. This peptide is predicted to be a major histocompatibility  
CC complex (MHC) binding peptide.  
XX  
SQ Sequence 15 AA;

Query Match 11.5%; Score 15; DB 22; Length 15;  
Best Local Similarity 100.0%; Pred. No. 1.5e-07;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 17 MSGPGQTSVAPPPE 31  
DB 1 MSGPGQTSVAPPPE 15

## RESULT 50

ABBI2859  
ID ABBI2859 standard; Peptide; 15 AA.

XX  
AC ABBI2859;

DT 22-JAN-2002 (first entry)

DE Human C35 peptide epitope #363.

XX  
XX Human; C35; cytostatic; gene therapy; vaccine; tumour antigen;  
KM breast cancer; bladder cancer; tumour immunotherapy; epitope;  
KM major histocompatibility complex binding peptide; MHC.

XX  
OS Homo sapiens.

XX  
PN WO200174859-A2.

XX  
PD 11-OCT-2001.

XX  
PF 04-APR-2001; 2001WO-US10855.

XX  
PR 04-APR-2000; 2000US-194463P.

XX  
PA (UYRP ) UNIV ROCHESTER.

XX  
PI Zauderer M, Evans EE, Borrello MA;

XX  
DR WPI; 2001-626383/72.

XX  
PT Novel C35 polypeptides and C35 genes useful in immunogenic compositions  
PT and vaccines, for inducing antibody and cell-mediated immunity against  
PT target cells, such as tumor cells that express C35 gene -  
XX  
PS Disclosure; Page 62; 33pp; English.

XX  
XX The present invention relates to human C35 (see AAG78997). C35 is a novel  
CC tumour antigen that is overexpressed in human breast and bladder  
CC carcinoma. C35 is thought to be a promising candidate for tumour  
CC immunotherapy, in immunogenic compositions and vaccines, to induce  
CC antibody and cell-mediated immunity against target cells such as tumour  
CC cells that express C35 genes. The present sequence is a C35 peptide  
CC epitope. This peptide is predicted to be a major histocompatibility  
CC complex (MHC) binding peptide.  
XX

SQ Sequence 15 AA;

Query Match 11.5%; Score 15; DB 22; Length 15;  
Best Local Similarity 100.0%; Pred. No. 1.5e-07;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 20 EPQTSVAPPPEVE 34  
DB 1 EPQTSVAPPPEVE 15

Search completed: December 15, 2003, 17:05:28  
Job time : 42 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

## OM protein - protein search, using sw model

Run on: December 15, 2003, 17:06:57 ; Search time 31 Seconds  
(without alignments)  
785,931 Million cell updates/sec

Title: US-09-925-301-966

Perfect score: 131  
Sequence: 1 AEVHTRKQPEAPPAAMSGE.....ASNGETLEKTNRPCCVIL 131

Scoring table: OLIGO  
Gapco 60.0 , Gapext 60.0

Searched: 684280 seqs, 185983659 residues

Word size : 0

Total number of hits satisfying chosen parameters: 684280

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Listing first 100 summaries

## Database :

Published Applications AA:\*

- 1: /cgn2\_6/ptodata/1/pubppa/US07\_PUBCOMB.pep:\*
- 2: /cgn2\_6/ptodata/1/pubppa/PCT\_NEW\_PUB.pep:\*
- 3: /cgn2\_6/ptodata/1/pubppa/US06\_NEW\_PUB.pep:\*
- 4: /cgn2\_6/ptodata/1/pubppa/US06\_PUBCOMB.pep:\*
- 5: /cgn2\_6/ptodata/1/pubppa/US07\_NEW\_PUB.pep:\*
- 6: /cgn2\_6/ptodata/1/pubppa/PCTUS\_PUBCOMB.pep:\*
- 7: /cgn2\_6/ptodata/1/pubppa/US08\_NEW\_PUB.pep:\*
- 8: /cgn2\_6/ptodata/1/pubppa/US09\_PUBCOMB.pep:\*
- 9: /cgn2\_6/ptodata/1/pubppa/US09\_PUBCOMB.pep:\*
- 10: /cgn2\_6/ptodata/1/pubppa/US09\_PUBCOMB.pep:\*
- 11: /cgn2\_6/ptodata/1/pubppa/US09\_PUBCOMB.pep:\*
- 12: /cgn2\_6/ptodata/1/pubppa/US09\_NEW\_PUB.pep:\*
- 13: /cgn2\_6/ptodata/1/pubppa/US10\_PUBCOMB.pep:\*
- 14: /cgn2\_6/ptodata/1/pubppa/US10\_PUBCOMB.pep:\*
- 15: /cgn2\_6/ptodata/1/pubppa/US10\_PUBCOMB.pep:\*
- 16: /cgn2\_6/ptodata/1/pubppa/US10\_NEW\_PUB.pep:\*
- 17: /cgn2\_6/ptodata/1/pubppa/US60\_NEW\_PUB.pep:\*
- 18: /cgn2\_6/ptodata/1/pubppa/US60\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	131	100.0	131	9	US-09-925-301-966
2	120	91.6	206	15	US-10-177-293-480
3	117	89.3	117	15	US-09-833-203-34
4	115	87.8	115	10	US-09-824-787B-2
5	10	7.6	10	12	US-09-833-203-41
6	10	7.6	10	12	US-09-833-203-44
7	10	7.6	10	12	US-09-833-203-46
8	10	7.6	10	12	US-09-833-203-50
9	10	7.6	10	12	US-09-833-203-56
10	10	7.6	10	12	US-09-833-203-58
11	10	7.6	10	12	US-09-833-203-60
12	9	6.9	9	12	US-09-833-203-38
13	9	6.9	9	12	US-09-833-203-43
14	9	6.9	9	12	US-09-833-203-47
15	9	6.9	9	12	US-09-833-203-49

16	9	6.9	9	12	US-09-833-203-51	Sequence 51, App1
17	9	6.9	9	12	US-09-833-203-53	Sequence 53, App1
18	9	6.9	9	12	US-09-833-203-54	Sequence 54, App1
19	9	6.9	9	12	US-09-833-203-59	Sequence 59, App1
20	9	6.9	9	12	US-09-833-203-61	Sequence 61, App1
21	9	6.9	9	12	US-09-833-203-62	Sequence 62, App1
22	9	6.9	9	12	US-09-833-203-63	Sequence 63, App1
23	9	6.9	9	12	US-09-824-787B-107	Sequence 107, App
24	9	6.9	9	10	US-09-824-787B-109	Sequence 109, App
25	9	6.9	9	10	US-09-824-787B-116	Sequence 116, App
26	9	6.9	9	10	US-09-824-787B-121	Sequence 121, App
27	9	6.9	9	10	US-09-824-787B-123	Sequence 123, App
28	9	6.9	9	10	US-09-824-787B-125	Sequence 125, App
29	9	6.9	9	10	US-09-824-787B-134	Sequence 134, App
30	9	6.9	9	10	US-09-824-787B-144	Sequence 144, App
31	9	6.9	9	10	US-09-824-787B-145	Sequence 145, App
32	9	6.9	9	10	US-09-824-787B-147	Sequence 147, App
33	9	6.1	8	12	US-09-833-203-39	Sequence 39, App1
34	8	6.1	8	12	US-09-833-203-40	Sequence 40, App1
35	8	6.1	8	12	US-09-833-203-42	Sequence 42, App1
36	8	6.1	8	12	US-09-833-203-45	Sequence 45, App1
37	8	6.1	8	12	US-09-833-203-48	Sequence 48, App1
38	8	6.1	8	12	US-09-833-203-52	Sequence 52, App1
39	8	6.1	8	12	US-09-833-203-55	Sequence 55, App1
40	8	6.1	8	12	US-09-833-203-57	Sequence 57, App1
41	8	6.1	9	10	US-09-824-787B-94	Sequence 94, App1
42	8	6.1	9	10	US-09-824-787B-96	Sequence 96, App1
43	8	6.1	9	10	US-09-824-787B-104	Sequence 104, App
44	8	6.1	9	10	US-09-824-787B-106	Sequence 106, App
45	8	6.1	9	10	US-09-824-787B-108	Sequence 108, App
46	8	6.1	9	10	US-09-824-787B-110	Sequence 110, App
47	8	6.1	9	10	US-09-824-787B-117	Sequence 117, App
48	8	6.1	9	10	US-09-824-787B-120	Sequence 120, App
49	8	6.1	9	10	US-09-824-787B-122	Sequence 122, App
50	8	6.1	9	10	US-09-824-787B-133	Sequence 133, App
51	8	6.1	9	10	US-09-824-787B-143	Sequence 143, App
52	8	6.1	10	10	US-09-824-787B-89	Sequence 89, App1
53	8	6.1	10	10	US-09-824-787B-93	Sequence 93, App1
54	8	6.1	10	10	US-09-824-787B-95	Sequence 95, App1
55	8	6.1	10	10	US-09-824-787B-98	Sequence 98, App1
56	8	6.1	10	10	US-09-824-787B-100	Sequence 100, App
57	8	6.1	10	10	US-09-824-787B-113	Sequence 113, App
58	8	6.1	10	10	US-09-824-787B-114	Sequence 114, App
59	8	6.1	10	10	US-09-824-787B-119	Sequence 119, App
60	8	6.1	10	10	US-09-824-787B-129	Sequence 129, App
61	8	6.1	10	10	US-09-824-787B-131	Sequence 131, App
62	8	6.1	10	10	US-09-824-787B-136	Sequence 136, App
63	8	6.1	10	10	US-09-824-787B-138	Sequence 138, App
64	8	6.1	10	10	US-09-824-787B-140	Sequence 140, App
65	8	6.1	10	10	US-09-824-787B-142	Sequence 142, App
66	8	5.3	8	10	US-09-824-787B-116	Sequence 116, App
67	7	5.3	9	10	US-09-824-787B-85	Sequence 85, App1
68	7	5.3	9	10	US-09-824-787B-91	Sequence 91, App1
69	7	5.3	9	10	US-09-824-787B-97	Sequence 97, App1
70	7	5.3	9	10	US-09-824-787B-99	Sequence 99, App1
71	7	5.3	9	10	US-09-824-787B-101	Sequence 101, App
72	7	5.3	9	10	US-09-824-787B-102	Sequence 102, App
73	7	5.3	9	10	US-09-824-787B-112	Sequence 112, App
74	7	5.3	9	10	US-09-824-787B-115	Sequence 115, App
75	7	5.3	9	10	US-09-824-787B-126	Sequence 126, App
76	7	5.3	9	10	US-09-824-787B-128	Sequence 128, App
77	7	5.3	9	10	US-09-824-787B-130	Sequence 130, App
78	7	5.3	9	10	US-09-824-787B-132	Sequence 132, App
79	7	5.3	9	10	US-09-824-787B-135	Sequence 135, App
80	7	5.3	10	10	US-09-824-787B-90	Sequence 90, App1
81	7	5.3	10	10	US-09-824-787B-105	Sequence 105, App
82	7	5.3	10	10	US-09-824-787B-111	Sequence 111, App
83	7	5.3	10	10	US-09-824-787B-127	Sequence 127, App
84	7	5.3	144	12	US-09-764-877-1904	Sequence 1904, App
85	7	5.3	164	12	US-10-094-749-1889	Sequence 1889, App
86	7	5.3	282	12	US-10-029-386-33647	Sequence 33647, App
87	7	5.3	445	9	US-09-815-242-11227	Sequence 11227, App
88	7	5.3	445	9	US-09-815-242-11243	Sequence 11243, App

89 7 5.3 520 15 US-10-156-761-14151 Sequence 14151, A  
90 7 5.3 622 12 US-09-738-626-4817 Sequence 4817, Ap  
91 7 5.3 628 10 US-10-145-479-4 Sequence 4, Appli  
92 7 5.3 749 15 US-10-211-962-100 Sequence 100, App  
93 7 5.3 924 12 US-10-145-479-2 Sequence 2, Appli  
94 7 5.3 1729 12 US-09-840-743-2 Sequence 103, App  
95 6 4.6 9 10 US-09-824-787B-103 Sequence 124, App  
96 6 4.6 9 10 US-09-824-787B-124 Sequence 137, App  
97 6 4.6 9 10 US-09-824-787B-137 Sequence 139, App  
98 6 4.6 9 10 US-09-824-787B-139 Sequence 141, App  
99 6 4.6 9 10 US-09-824-787B-141 Sequence 146, App  
100 6 4.6 9 10 US-09-824-787B-146

## ALIGNMENTS

RESULT 1  
US-09-925-301-966  
Sequence 966, Application US/09925301  
Patent No. US20020052308A1  
GENERAL INFORMATION:  
APPLICANT: Rosen et al.  
TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies  
FILE REFERENCE: PA106  
CURRENT APPLICATION NUMBER: US/09/925,301  
CURRENT FILING DATE: 2001-08-10  
PRIOR APPLICATION NUMBER: PCT/US00/05882  
PRIOR FILING DATE: 2000-03-08  
PRIOR APPLICATION NUMBER: 60/124,270  
PRIOR FILING DATE: 1999-03-12  
NUMBER OF SEQ ID NOS: 1694  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 966  
LENGTH: 131  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-925-301-966

Query Match 100.0%; Score 131; DB 9; Length 131;  
Best Local Similarity 100.0%; Pred. No. 4,5e-123;  
Matches 131; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AEVHTROGPEAPPAAMSGERGQTSVAPPEEVEPGSGVRIVVEYCEPCGFEATYLELAS 60  
DB 1 AEVHTROGPEAPPAAMSGERGQTSVAPPEEVEPGSGVRIVVEYCEPCGFEATYLELAS 60  
QY 61 AVEQYVGIIEISRLGCTGAFEIRINQVLVPSKLENGFPYKDLIAIRASNGETLEK 120  
DB 61 AVEQYVGIIEISRLGCTGAFEIRINQVLVPSKLENGFPYKDLIAIRASNGETLEK 120  
QY 121 ITNSRPPCVIL 131  
DB 121 ITNSRPPCVIL 131  
RESULT 2  
US-10-177-293-480  
Sequence 480, Application US/10177293  
Publication No. US20030124128A1  
GENERAL INFORMATION:  
APPLICANT: Lillie, James  
APPLICANT: Glatt, Karen  
APPLICANT: Zhao, Xumei  
APPLICANT: Ganmavaypu, Manjula  
APPLICANT: Kamatkar, Shubhangi  
APPLICANT: Mertens, Maureen  
APPLICANT: Meyer, Vic  
APPLICANT: Wang, Youzhen  
APPLICANT: Xu, Yongyao  
APPLICANT: Hoersch, Sebastian  
APPLICANT: Monahan, John  
APPLICANT: Meyers, Rachel E.

APPLICANT: Baot Jr., Robert C.  
APPLICANT: Hortobagyi, Gabriel N.  
APPLICANT: Pusztai, Lajos  
APPLICANT: Meric, Funda  
APPLICANT: Sahin, Aysegul  
APPLICANT: Mills, Gordon B.  
TITLE OF INVENTION: COMPOSITIONS, KITS, AND METHODS FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY OF BREAST CANCER  
FILE REFERENCE: MRI-038  
CURRENT APPLICATION NUMBER: US/10/177,293  
CURRENT FILING DATE: 2002-06-21  
PRIOR APPLICATION NUMBER: US 60/299,887  
PRIOR FILING DATE: 2001-06-21  
PRIOR APPLICATION NUMBER: US 60/301,572  
PRIOR FILING DATE: 2001-06-27  
PRIOR APPLICATION NUMBER: US 60/306,501  
PRIOR FILING DATE: 2001-07-18  
PRIOR APPLICATION NUMBER: US 60/325,002  
PRIOR FILING DATE: 2001-09-25  
PRIOR APPLICATION NUMBER: US 60/362,585  
PRIOR FILING DATE: 2002-03-05  
PRIOR APPLICATION NUMBER: US 60/xxx,xxx  
PRIOR FILING DATE: 2002-05-14  
NUMBER OF SEQ ID NOS: 506  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 480  
LENGTH: 206  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-177-293-480

Query Match 91.6%; Score 120; DB 15; Length 206;  
Best Local Similarity 100.0%; Pred. No. 6.7e-112;  
Matches 120; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 AEPAMSGERGQTSVAPPEEVEPGSGVRIVVEYCEPCGFEATYLELASAVEQYVGIIEI 71  
DB 87 AEPAMSGERGQTSVAPPEEVEPGSGVRIVVEYCEPCGFEATYLELASAVEQYVGIIEI 146  
QY 72 ESRIGCTGAFEIRINQVLVPSKLENGFPYKDLIAIRASNGETLEKITSRPPCVIL 131  
DB 147 ESRIGCTGAFEIRINQVLVPSKLENGFPYKDLIAIRASNGETLEKITSRPPCVIL 206

RESULT 3  
US-09-833-203-34  
Sequence 34, Application US/09833203  
Publication No. US2003016277A1  
GENERAL INFORMATION:  
APPLICANT: Zauderer, Maurice  
APPLICANT: Smith, Ernest S.  
TITLE OF INVENTION: Targeted Vaccine Delivery Systems  
FILE REFERENCE: 1821.0020001  
CURRENT APPLICATION NUMBER: US/09/833,203  
CURRENT FILING DATE: 2001-04-12  
PRIOR APPLICATION NUMBER: US 60/196,472  
PRIOR FILING DATE: 2000-04-12  
NUMBER OF SEQ ID NOS: 63  
SOFTWARE: PatentIn Version 3.0  
SEQ ID NO 34  
LENGTH: 117  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
NAME/KEY: misc\_feature  
OTHER INFORMATION: C35  
US-09-833-203-34

Query Match 89.3%; Score 117; DB 12; Length 117;  
Best Local Similarity 100.0%; Pred. No. 4e-109;  
Matches 117; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 15 AAMSGERGQTSVAPPEEVEPGSGVRIVVEYCEPCGFEATYLELASAVEQYVGIIEISR 74

Db 1 AAMSGPGQTSVAPPEEVEPGSGVRIVVEYCEPGFEATYIELASAVEQYPGIEISR 60  
Qy 75 LGGTGAFEIINGQLVFSKLNGGFPYEKDLIEAIRRANGETLEKITSRPPCVIL 131  
Db 61 LGGTGAFEIINGQLVFSKLNGGFPYEKDLIEAIRRANGETLEKITSRPPCVIL 117

## RESULT 4

US-09-824-787B-2  
; Sequence 2, Application US/09824787B  
; Patent No. US20020155447A1  
; GENERAL INFORMATION:  
; APPLICANT: Zauderer, Maurice  
; APPLICANT: Evans, Elizabeth E.  
; APPLICANT: Borrello, Melinda A.  
; TITLE OF INVENTION: A Gene differentially Expressed in Breast and  
; TITLE OF INVENTION: Bladder Cancer, and Encoded Polypeptides  
; FILE REFERENCE: 1821.004001  
; CURRENT APPLICATION NUMBER: US/09/824,787B  
; CURRENT FILING DATE: 2001-04-04  
; PRIOR APPLICATION NUMBER: 60/194,463  
; PRIOR FILING DATE: 2000-04-04  
; NUMBER OF SEQ ID NOS: 147  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 2  
; LENGTH: 115  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-824-787B-2

Query Match 87.8%; Score 115; DB 10; Length 115;  
Best Local Similarity 100.0%; Pred. No. 4e-107;  
Matches 115; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 17 MSGERGQTSVAPPEEVEPGSGVRIVVEYCEPGFEATYIELASAVEQYPGIEISR 76  
Db 1 MSGERGQTSVAPPEEVEPGSGVRIVVEYCEPGFEATYIELASAVEQYPGIEISR 60  
Qy 77 GTGAFEIINGQLVFSKLNGGFPYEKDLIEAIRRANGETLEKITSRPPCVIL 131  
Db 61 GTGAFEIINGQLVFSKLNGGFPYEKDLIEAIRRANGETLEKITSRPPCVIL 115

## RESULT 5

US-09-833-203-41  
; Sequence 41, Application US/09833203  
; Publication No. US20030166277A1  
; GENERAL INFORMATION:  
; APPLICANT: Zauderer, Maurice  
; APPLICANT: Smith, Ernest S.  
; TITLE OF INVENTION: Targeted Vaccine Delivery Systems  
; FILE REFERENCE: 1821.002001  
; CURRENT APPLICATION NUMBER: US/09/833,203  
; CURRENT FILING DATE: 2001-04-12  
; PRIOR APPLICATION NUMBER: US 60/196,472  
; PRIOR FILING DATE: 2000-04-12  
; NUMBER OF SEQ ID NOS: 63  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 41  
; LENGTH: 10  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; NAME/KEY: misc feature  
; OTHER INFORMATION: C35 peptides  
US-09-833-203-41

Query Match 7.6%; Score 10; DB 12; Length 10;  
Best Local Similarity 100.0%; Pred. No. 0.004;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 32 EEPGSGVRI 41

Db 1 EEPGSGVRI 10

RESULT 6  
US-09-833-203-44  
; Sequence 44, Application US/09833203  
; Publication No. US20030166277A1  
; GENERAL INFORMATION:  
; APPLICANT: Zauderer, Maurice  
; APPLICANT: Smith, Ernest S.  
; TITLE OF INVENTION: Targeted Vaccine Delivery Systems  
; FILE REFERENCE: 1821.002001  
; CURRENT APPLICATION NUMBER: US/09/833,203  
; CURRENT FILING DATE: 2001-04-12  
; PRIOR APPLICATION NUMBER: US 60/196,472  
; PRIOR FILING DATE: 2000-04-12  
; NUMBER OF SEQ ID NOS: 63  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 44  
; LENGTH: 10  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; NAME/KEY: misc feature  
; OTHER INFORMATION: C35 peptides  
US-09-833-203-44

Query Match 7.6%; Score 10; DB 12; Length 10;  
Best Local Similarity 100.0%; Pred. No. 0.004;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 53 ATYLEIASAV 62  
Db 1 ATYLEIASAV 10  
RESULT 7  
US-09-833-203-46  
; Sequence 46, Application US/09833203  
; Publication No. US20030166277A1  
; GENERAL INFORMATION:  
; APPLICANT: Zauderer, Maurice  
; APPLICANT: Smith, Ernest S.  
; TITLE OF INVENTION: Targeted Vaccine Delivery Systems  
; FILE REFERENCE: 1821.002001  
; CURRENT APPLICATION NUMBER: US/09/833,203  
; CURRENT FILING DATE: 2001-04-12  
; PRIOR APPLICATION NUMBER: US 60/196,472  
; PRIOR FILING DATE: 2000-04-12  
; NUMBER OF SEQ ID NOS: 63  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 46  
; LENGTH: 10  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; NAME/KEY: misc feature  
; OTHER INFORMATION: C35 peptides  
US-09-833-203-46

Query Match 7.6%; Score 10; DB 12; Length 10;  
Best Local Similarity 100.0%; Pred. No. 0.004;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 60 SAVKEQYPGI 69  
Db 1 SAVKEQYPGI 10

RESULT 8  
US-09-833-203-50  
; Sequence 50, Application US/09833203

```

; Publication No. US20030166277A1
; GENERAL INFORMATION:
; APPLICANT: Zauderer, Maurice
; APPLICANT: Smith, Ernest S.
; TITLE OF INVENTION: Targeted Vaccine Delivery Systems
; FILE REFERENCE: 1821.0020001
; CURRENT APPLICATION NUMBER: US/09/833,203
; CURRENT FILING DATE: 2001-04-12
; PRIOR APPLICATION NUMBER: US 60/196,472
; PRIOR FILING DATE: 2000-04-12
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 50
; LENGTH: 10
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: C35 peptides
US-09-833-203-50

```

```

Query Match      7.6%; Score 10; DB 12; Length 10;
Best Local Similarity 100.0%; Pred. No. 0.004;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      74 RLGGTGAFFI 83
DB      1 RLGGTGAFFI 10

```

```

RESULT 9
US-09-833-203-56
; Sequence 56, Application US/09833203
; Publication No. US20030166277A1
; GENERAL INFORMATION:
; APPLICANT: Zauderer, Maurice
; APPLICANT: Smith, Ernest S.
; TITLE OF INVENTION: Targeted Vaccine Delivery Systems
; FILE REFERENCE: 1821.0020001
; CURRENT APPLICATION NUMBER: US/09/833,203
; CURRENT FILING DATE: 2001-04-12
; PRIOR APPLICATION NUMBER: US 60/196,472
; PRIOR FILING DATE: 2000-04-12
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 56
; LENGTH: 10
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: C35 peptides
US-09-833-203-56

```

```

Query Match      7.6%; Score 10; DB 12; Length 10;
Best Local Similarity 100.0%; Pred. No. 0.004;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      108 AIRRASNGET 117
DB      1 AIRRASNGET 10

```

```

RESULT 10
US-09-833-203-58
; Sequence 58, Application US/09833203
; Publication No. US20030166277A1
; GENERAL INFORMATION:
; APPLICANT: Zauderer, Maurice
; APPLICANT: Smith, Ernest S.
; TITLE OF INVENTION: Targeted Vaccine Delivery Systems
; FILE REFERENCE: 1821.0020001
; CURRENT APPLICATION NUMBER: US/09/833,203

```

```

; CURRENT FILING DATE: 2001-04-12
; PRIOR APPLICATION NUMBER: US 60/196,472
; PRIOR FILING DATE: 2000-04-12
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 58
; LENGTH: 10
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: C35 peptides
US-09-833-203-58

```

```

Query Match      7.6%; Score 10; DB 12; Length 10;
Best Local Similarity 100.0%; Pred. No. 0.004;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      120 KITNSRPPCV 129
DB      1 KITNSRPPCV 10

```

```

RESULT 11
US-09-833-203-60
; Sequence 60, Application US/09833203
; Publication No. US20030166277A1
; GENERAL INFORMATION:
; APPLICANT: Zauderer, Maurice
; APPLICANT: Smith, Ernest S.
; TITLE OF INVENTION: Targeted Vaccine Delivery Systems
; FILE REFERENCE: 1821.0020001
; CURRENT APPLICATION NUMBER: US/09/833,203
; CURRENT FILING DATE: 2001-04-12
; PRIOR APPLICATION NUMBER: US 60/196,472
; PRIOR FILING DATE: 2000-04-12
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 60
; LENGTH: 10
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: C35 peptides
US-09-833-203-60

```

```

Query Match      7.6%; Score 10; DB 12; Length 10;
Best Local Similarity 100.0%; Pred. No. 0.004;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      121 ITNSRPPCVI 130
DB      1 ITNSRPPCVI 10

```

```

RESULT 12
US-09-833-203-38
; Sequence 38, Application US/09833203
; Publication No. US20030166277A1
; GENERAL INFORMATION:
; APPLICANT: Zauderer, Maurice
; APPLICANT: Smith, Ernest S.
; TITLE OF INVENTION: Targeted Vaccine Delivery Systems
; FILE REFERENCE: 1821.0020001
; CURRENT APPLICATION NUMBER: US/09/833,203
; CURRENT FILING DATE: 2001-04-12
; PRIOR APPLICATION NUMBER: US 60/196,472
; PRIOR FILING DATE: 2000-04-12
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 38
; LENGTH: 9

```

TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
NAME/KEY: misc feature  
OTHER INFORMATION: C35 peptides  
US-09-833-203-38

Query Match  
Best Local Similarity 100.0%; Score 9; DB 12; Length 9;  
Pred. No. 6.2e+05;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 25 SVAPPEEV 33  
DB 1 SVAPPEEV 9

RESULT 13  
US-09-833-203-43

Sequence 43, Application US/09833203  
Publication No. US2003016277A1  
GENERAL INFORMATION:  
APPLICANT: Zauderer, Maurice  
APPLICANT: Smith, Ernest S.  
TITLE OF INVENTION: Targeted Vaccine Delivery Systems  
FILE REFERENCE: 1821.0020001  
CURRENT APPLICATION NUMBER: US/09/833,203  
CURRENT FILING DATE: 2001-04-12  
PRIOR APPLICATION NUMBER: US 60/196,472  
PRIOR FILING DATE: 2000-04-12  
NUMBER OF SEQ ID NOS: 63  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 43  
LENGTH: 9  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
NAME/KEY: misc feature  
OTHER INFORMATION: C35 peptides  
US-09-833-203-43

Query Match  
Best Local Similarity 100.0%; Score 9; DB 12; Length 9;  
Pred. No. 6.2e+05;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 53 ATYLELASA 61  
DB 1 ATYLELASA 9

RESULT 14  
US-09-833-203-47

Sequence 47, Application US/09833203  
Publication No. US2003016277A1  
GENERAL INFORMATION:  
APPLICANT: Zauderer, Maurice  
APPLICANT: Smith, Ernest S.  
TITLE OF INVENTION: Targeted Vaccine Delivery Systems  
FILE REFERENCE: 1821.0020001  
CURRENT APPLICATION NUMBER: US/09/833,203  
CURRENT FILING DATE: 2001-04-12  
PRIOR APPLICATION NUMBER: US 60/196,472  
PRIOR FILING DATE: 2000-04-12  
NUMBER OF SEQ ID NOS: 63  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 47  
LENGTH: 9  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
NAME/KEY: misc feature  
OTHER INFORMATION: C35 peptides  
US-09-833-203-47

Query Match  
Best Local Similarity 100.0%; Score 9; DB 12; Length 9;  
Pred. No. 6.2e+05;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 61 AVEQYRPGI 69  
DB 1 AVEQYRPGI 9

RESULT 15  
US-09-833-203-49

Sequence 49, Application US/09833203  
Publication No. US2003016277A1  
GENERAL INFORMATION:  
APPLICANT: Zauderer, Maurice  
APPLICANT: Smith, Ernest S.  
TITLE OF INVENTION: Targeted Vaccine Delivery Systems  
FILE REFERENCE: 1821.0020001  
CURRENT APPLICATION NUMBER: US/09/833,203  
CURRENT FILING DATE: 2001-04-12  
PRIOR APPLICATION NUMBER: US 60/196,472  
PRIOR FILING DATE: 2000-04-12  
NUMBER OF SEQ ID NOS: 63  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 49  
LENGTH: 9  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
NAME/KEY: misc feature  
OTHER INFORMATION: C35 peptides  
US-09-833-203-49

Query Match  
Best Local Similarity 100.0%; Score 9; DB 12; Length 9;  
Pred. No. 6.2e+05;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 70 EIESRLGGT 78  
DB 1 EIESRLGGT 9

RESULT 16  
US-09-833-203-51

Sequence 51, Application US/09833203  
Publication No. US2003016277A1  
GENERAL INFORMATION:  
APPLICANT: Zauderer, Maurice  
APPLICANT: Smith, Ernest S.  
TITLE OF INVENTION: Targeted Vaccine Delivery Systems  
FILE REFERENCE: 1821.0020001  
CURRENT APPLICATION NUMBER: US/09/833,203  
CURRENT FILING DATE: 2001-04-12  
PRIOR APPLICATION NUMBER: US 60/196,472  
PRIOR FILING DATE: 2000-04-12  
NUMBER OF SEQ ID NOS: 63  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 51  
LENGTH: 9  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
NAME/KEY: misc feature  
OTHER INFORMATION: C35 peptides  
US-09-833-203-51

Query Match  
Best Local Similarity 100.0%; Score 9; DB 12; Length 9;  
Pred. No. 6.2e+05;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 77 GTGAFIEI 85  
DB 1 GTGAFIEI 9

```
RESULT 17
US-09-833-203-53
; Sequence 53, Application US/09833203
; Publication No. US20030166277A1
; GENERAL INFORMATION:
; APPLICANT: Zauderer, Maurice
; TITLE OF INVENTION: Targeted Vaccine Delivery Systems
; FILE REFERENCE: 1821.0020001
; CURRENT APPLICATION NUMBER: US/09/833,203
; CURRENT FILING DATE: 2001-04-12
; PRIOR APPLICATION NUMBER: US 60/196,472
; PRIOR FILING DATE: 2000-04-12
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 53
; LENGTH: 9
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: C35 peptides
US-09-833-203-53

Query Match
Best Local Similarity 100.0%; Score 9; DB 12; Length 9;
Pred. No. 6.2e+05;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 82 EIEINGQLV 90
Db 1 EIEINGQLV 9

RESULT 18
US-09-833-203-54
; Sequence 54, Application US/09833203
; Publication No. US20030166277A1
; GENERAL INFORMATION:
; APPLICANT: Zauderer, Maurice
; TITLE OF INVENTION: Targeted Vaccine Delivery Systems
; FILE REFERENCE: 1821.0020001
; CURRENT APPLICATION NUMBER: US/09/833,203
; CURRENT FILING DATE: 2001-04-12
; PRIOR APPLICATION NUMBER: US 60/196,472
; PRIOR FILING DATE: 2000-04-12
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 54
; LENGTH: 9
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: C35 peptides
US-09-833-203-54

Query Match
Best Local Similarity 100.0%; Score 9; DB 12; Length 9;
Pred. No. 6.2e+05;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 104 DLIEAIRRA 112
Db 1 DLIEAIRRA 9

RESULT 19
US-09-833-203-59
; Sequence 59, Application US/09833203
; Publication No. US20030166277A1
; GENERAL INFORMATION:
```

```
; APPLICANT: Zauderer, Maurice
; TITLE OF INVENTION: Targeted Vaccine Delivery Systems
; FILE REFERENCE: 1821.0020001
; CURRENT APPLICATION NUMBER: US/09/833,203
; CURRENT FILING DATE: 2001-04-12
; PRIOR APPLICATION NUMBER: US 60/196,472
; PRIOR FILING DATE: 2000-04-12
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 59
; LENGTH: 9
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: C35 peptides
US-09-833-203-59

Query Match
Best Local Similarity 100.0%; Score 9; DB 12; Length 9;
Pred. No. 6.2e+05;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 121 ITNSRPPCV 129
Db 1 ITNSRPPCV 9

RESULT 20
US-09-833-203-61
; Sequence 61, Application US/09833203
; Publication No. US20030166277A1
; GENERAL INFORMATION:
; APPLICANT: Zauderer, Maurice
; TITLE OF INVENTION: Targeted Vaccine Delivery Systems
; FILE REFERENCE: 1821.0020001
; CURRENT APPLICATION NUMBER: US/09/833,203
; CURRENT FILING DATE: 2001-04-12
; PRIOR APPLICATION NUMBER: US 60/196,472
; PRIOR FILING DATE: 2000-04-12
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 61
; LENGTH: 9
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: C35 peptides
US-09-833-203-61

Query Match
Best Local Similarity 100.0%; Score 9; DB 12; Length 9;
Pred. No. 6.2e+05;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 32 EVERGSGVR 40
Db 1 EVERGSGVR 9

RESULT 21
US-09-833-203-62
; Sequence 62, Application US/09833203
; Publication No. US20030166277A1
; GENERAL INFORMATION:
; APPLICANT: Zauderer, Maurice
; TITLE OF INVENTION: Targeted Vaccine Delivery Systems
; FILE REFERENCE: 1821.0020001
; CURRENT APPLICATION NUMBER: US/09/833,203
; CURRENT FILING DATE: 2001-04-12
; PRIOR APPLICATION NUMBER: US 60/196,472
```



PRIOR FILING DATE: 2000-04-12  
NUMBER OF SEQ ID NOS: 63  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 62  
LENGTH: 9  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
NAME/KEY: misc feature  
OTHER INFORMATION: C35 peptides  
US-09-833-203-62

Query Match  
Best Local Similarity 100.0%; Score 9; DB 12; Length 9;  
Pred. No. 6.2e+05;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 47 EPCGFATY 55  
Db 1 EPCGFATY 9

RESULT 22  
US-09-833-203-63  
Sequence 63, Application US/09833203  
Publication No. US20030166277A1  
GENERAL INFORMATION:  
APPLICANT: Zauderer, Maurice  
APPLICANT: Smith, Ernest S.  
TITLE OF INVENTION: Targeted Vaccine Delivery Systems  
FILE REFERENCE: 1821.0020001  
CURRENT FILING DATE: 2001-04-12  
PRIOR APPLICATION NUMBER: US/09/833,203  
PRIOR FILING DATE: 2000-04-12  
NUMBER OF SEQ ID NOS: 63  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 63  
LENGTH: 9  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
NAME/KEY: misc feature  
OTHER INFORMATION: C35 peptides  
US-09-833-203-63

Query Match  
Best Local Similarity 100.0%; Score 9; DB 12; Length 9;  
Pred. No. 6.2e+05;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 112 ASNGETLEK 120  
Db 1 ASNGETLEK 9

RESULT 23  
US-09-824-787B-107  
Sequence 107, Application US/09824787B  
Patent No. US20020155447A1  
GENERAL INFORMATION:  
APPLICANT: Zauderer, Maurice  
APPLICANT: Evans, Elizabeth E.  
APPLICANT: Borrello, Melinda A.  
TITLE OF INVENTION: A Gene Differentially Expressed in Breast and  
FILE REFERENCE: 1821.0040001  
CURRENT APPLICATION NUMBER: US/09/824,787B  
CURRENT FILING DATE: 2001-04-04  
PRIOR APPLICATION NUMBER: 60/194,463  
PRIOR FILING DATE: 2000-04-04  
NUMBER OF SEQ ID NOS: 147  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 107  
LENGTH: 10

TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-824-787B-107

Query Match  
Best Local Similarity 100.0%; Score 9; DB 10; Length 10;  
Pred. No. 0.04;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 69 IEIESRLGG 77  
Db 1 IEIESRLGG 9

RESULT 24  
US-09-824-787B-109  
Sequence 109, Application US/09824787B  
Patent No. US20020155447A1  
GENERAL INFORMATION:  
APPLICANT: Zauderer, Maurice  
APPLICANT: Evans, Elizabeth E.  
APPLICANT: Borrello, Melinda A.  
TITLE OF INVENTION: A Gene Differentially Expressed in Breast and  
FILE REFERENCE: 1821.0040001  
CURRENT APPLICATION NUMBER: US/09/824,787B  
CURRENT FILING DATE: 2001-04-04  
PRIOR APPLICATION NUMBER: 60/194,463  
PRIOR FILING DATE: 2000-04-04  
NUMBER OF SEQ ID NOS: 147  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 109  
LENGTH: 10  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-824-787B-109

Query Match  
Best Local Similarity 100.0%; Score 9; DB 10; Length 10;  
Pred. No. 0.04;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 81 FEIEINGQL 89  
Db 1 FEIEINGQL 9

RESULT 25  
US-09-824-787B-116  
Sequence 116, Application US/09824787B  
Patent No. US20020155447A1  
GENERAL INFORMATION:  
APPLICANT: Zauderer, Maurice  
APPLICANT: Evans, Elizabeth E.  
APPLICANT: Borrello, Melinda A.  
TITLE OF INVENTION: A Gene Differentially Expressed in Breast and  
FILE REFERENCE: 1821.0040001  
CURRENT APPLICATION NUMBER: US/09/824,787B  
CURRENT FILING DATE: 2001-04-04  
PRIOR APPLICATION NUMBER: 60/194,463  
PRIOR FILING DATE: 2000-04-04  
NUMBER OF SEQ ID NOS: 147  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 116  
LENGTH: 10  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-824-787B-116

Query Match  
Best Local Similarity 100.0%; Score 9; DB 10; Length 10;  
Pred. No. 0.04;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 34 EPGSGVRIV 42

Db 1 EFGSGVIV 9

## RESULT 26

US-09-824-787B-121  
 ; Sequence 121, Application US/09824787B  
 ; Patent No. US20020155447A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Zauderer, Maurice  
 ; APPLICANT: Evans, Elizabeth E.  
 ; APPLICANT: Borrello, Melinda A.  
 ; TITLE OF INVENTION: A Gene Differentially Expressed in Breast and  
 ; TITLE OF INVENTION: Bladder Cancer, and Encoded Polypeptides  
 ; FILE REFERENCE: 1821.0040001  
 ; CURRENT APPLICATION NUMBER: US/09/824,787B  
 ; CURRENT FILING DATE: 2001-04-04  
 ; PRIOR APPLICATION NUMBER: 60/194,463  
 ; PRIOR FILING DATE: 2000-04-04  
 ; NUMBER OF SEQ ID NOS: 147  
 ; SOFTWARE: PatentIn Ver. 2.1  
 ; SEQ ID NO 121  
 ; LENGTH: 10  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 US-09-824-787B-121

Query Match 6.9%; Score 9; DB 10; Length 10;  
 Best Local Similarity 100.0%; Pred. No. 0.04;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 39 VRIIVEYCE 47  
 Db 1 VRIIVEYCE 9

## RESULT 27

US-09-824-787B-123  
 ; Sequence 123, Application US/09824787B  
 ; Patent No. US20020155447A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Zauderer, Maurice  
 ; APPLICANT: Evans, Elizabeth E.  
 ; APPLICANT: Borrello, Melinda A.  
 ; TITLE OF INVENTION: A Gene Differentially Expressed in Breast and  
 ; TITLE OF INVENTION: Bladder Cancer, and Encoded Polypeptides  
 ; FILE REFERENCE: 1821.0040001  
 ; CURRENT APPLICATION NUMBER: US/09/824,787B  
 ; CURRENT FILING DATE: 2001-04-04  
 ; PRIOR APPLICATION NUMBER: 60/194,463  
 ; PRIOR FILING DATE: 2000-04-04  
 ; NUMBER OF SEQ ID NOS: 147  
 ; SOFTWARE: PatentIn Ver. 2.1  
 ; SEQ ID NO 123  
 ; LENGTH: 10  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 US-09-824-787B-123

Query Match 6.9%; Score 9; DB 10; Length 10;  
 Best Local Similarity 100.0%; Pred. No. 0.04;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 110 RRASNGETL 118  
 Db 1 RRASNGETL 9

RESULT 28  
 US-09-824-787B-125  
 ; Sequence 125, Application US/09824787B  
 ; Patent No. US20020155447A1  
 ; GENERAL INFORMATION:

; APPLICANT: Zauderer, Maurice  
 ; APPLICANT: Evans, Elizabeth E.  
 ; APPLICANT: Borrello, Melinda A.  
 ; TITLE OF INVENTION: A Gene Differentially Expressed in Breast and  
 ; TITLE OF INVENTION: Bladder Cancer, and Encoded Polypeptides  
 ; FILE REFERENCE: 1821.0040001  
 ; CURRENT APPLICATION NUMBER: US/09/824,787B  
 ; CURRENT FILING DATE: 2001-04-04  
 ; PRIOR APPLICATION NUMBER: 60/194,463  
 ; PRIOR FILING DATE: 2000-04-04  
 ; NUMBER OF SEQ ID NOS: 147  
 ; SOFTWARE: PatentIn Ver. 2.1  
 ; SEQ ID NO 125  
 ; LENGTH: 10  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 US-09-824-787B-125

Query Match 6.9%; Score 9; DB 10; Length 10;  
 Best Local Similarity 100.0%; Pred. No. 0.04;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 99 FPEKDLIE 107  
 Db 1 FPEKDLIE 9

RESULT 29  
 US-09-824-787B-134  
 ; Sequence 134, Application US/09824787B  
 ; Patent No. US20020155447A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Zauderer, Maurice  
 ; APPLICANT: Evans, Elizabeth E.  
 ; APPLICANT: Borrello, Melinda A.  
 ; TITLE OF INVENTION: A Gene Differentially Expressed in Breast and  
 ; TITLE OF INVENTION: Bladder Cancer, and Encoded Polypeptides  
 ; FILE REFERENCE: 1821.0040001  
 ; CURRENT APPLICATION NUMBER: US/09/824,787B  
 ; CURRENT FILING DATE: 2001-04-04  
 ; PRIOR APPLICATION NUMBER: 60/194,463  
 ; PRIOR FILING DATE: 2000-04-04  
 ; NUMBER OF SEQ ID NOS: 147  
 ; SOFTWARE: PatentIn Ver. 2.1  
 ; SEQ ID NO 134  
 ; LENGTH: 10  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 US-09-824-787B-134

Query Match 6.9%; Score 9; DB 10; Length 10;  
 Best Local Similarity 100.0%; Pred. No. 0.04;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 69 IEIESRLGG 77  
 Db 1 IEIESRLGG 9

RESULT 30  
 US-09-824-787B-144  
 ; Sequence 144, Application US/09824787B  
 ; Patent No. US20020155447A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Zauderer, Maurice  
 ; APPLICANT: Evans, Elizabeth E.  
 ; APPLICANT: Borrello, Melinda A.  
 ; TITLE OF INVENTION: A Gene Differentially Expressed in Breast and  
 ; TITLE OF INVENTION: Bladder Cancer, and Encoded Polypeptides  
 ; FILE REFERENCE: 1821.0040001  
 ; CURRENT APPLICATION NUMBER: US/09/824,787B  
 ; CURRENT FILING DATE: 2001-04-04  
 ; PRIOR APPLICATION NUMBER: 60/194,463

;; PRIOR FILING DATE: 2000-04-04  
;; NUMBER OF SEQ ID NOS: 147  
;; SOFTWARE: PatentIn Ver. 2.1  
;; SEQ ID NO 144  
;; LENGTH: 10  
;; TYPE: PRT  
;; ORGANISM: Homo sapiens  
US-09-824-787B-144

Query Match 6.9%; Score 9; DB 10; Length 10;  
Best Local Similarity 100.0%; Pred. No. 0.04;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 36 GSGVRIVE 44  
Db 1 GSGVRIVE 9

RESULT 31  
US-09-824-787B-145  
;; Sequence 145, Application US/09824787B  
;; Patent No. US20020155447A1  
;; GENERAL INFORMATION:  
;; APPLICANT: Zauderer, Maurice  
;; APPLICANT: Evans, Elizabeth E.  
;; APPLICANT: Borrello, Melinda A.  
;; TITLE OF INVENTION: A Gene Differentially Expressed in Breast and  
;; TITLE OF INVENTION: Bladder Cancer, and Encoded Polypeptides  
;; FILE REFERENCE: 1821.0040001  
;; CURRENT APPLICATION NUMBER: US/09/824,787B  
;; PRIOR FILING DATE: 2001-04-04  
;; PRIOR APPLICATION NUMBER: 60/194,463  
;; NUMBER OF SEQ ID NOS: 147  
;; SOFTWARE: PatentIn Ver. 2.1  
;; SEQ ID NO 145  
;; LENGTH: 10  
;; TYPE: PRT  
;; ORGANISM: Homo sapiens  
US-09-824-787B-145

Query Match 6.9%; Score 9; DB 10; Length 10;  
Best Local Similarity 100.0%; Pred. No. 0.04;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 60 SAVKEQYPG 68  
Db 1 SAVKEQYPG 9

RESULT 32  
US-09-824-787B-147  
;; Sequence 147, Application US/09824787B  
;; Patent No. US20020155447A1  
;; GENERAL INFORMATION:  
;; APPLICANT: Zauderer, Maurice  
;; APPLICANT: Evans, Elizabeth E.  
;; APPLICANT: Borrello, Melinda A.  
;; TITLE OF INVENTION: A Gene Differentially Expressed in Breast and  
;; TITLE OF INVENTION: Bladder Cancer, and Encoded Polypeptides  
;; FILE REFERENCE: 1821.0040001  
;; CURRENT APPLICATION NUMBER: US/09/824,787B  
;; PRIOR FILING DATE: 2001-04-04  
;; PRIOR APPLICATION NUMBER: 60/194,463  
;; PRIOR FILING DATE: 2000-04-04  
;; NUMBER OF SEQ ID NOS: 147  
;; SOFTWARE: PatentIn Ver. 2.1  
;; SEQ ID NO 147  
;; LENGTH: 10  
;; TYPE: PRT  
;; ORGANISM: Homo sapiens  
US-09-824-787B-147

Query Match 6.9%; Score 9; DB 10; Length 10;  
Best Local Similarity 100.0%; Pred. No. 0.04;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 90 VFSEKLENGC 98  
Db 1 VFSEKLENGC 9

RESULT 33  
US-09-833-203-39  
;; Sequence 39, Application US/09833203  
;; Publication No. US20030166277A1  
;; GENERAL INFORMATION:  
;; APPLICANT: Zauderer, Maurice  
;; APPLICANT: Smith, Ernest S.  
;; TITLE OF INVENTION: Targeted Vaccine Delivery Systems  
;; FILE REFERENCE: 1821.0020001  
;; CURRENT APPLICATION NUMBER: US/09/833,203  
;; PRIOR FILING DATE: 2001-04-12  
;; PRIOR APPLICATION NUMBER: US 60/196,472  
;; PRIOR FILING DATE: 2000-04-12  
;; NUMBER OF SEQ ID NOS: 63  
;; SOFTWARE: PatentIn version 3.0  
;; SEQ ID NO 39  
;; LENGTH: 8  
;; TYPE: PRT  
;; ORGANISM: Artificial Sequence  
;; FEATURE:  
;; NAME/KEY: misc feature  
;; OTHER INFORMATION: C35 peptides  
US-09-833-203-39

Query Match 6.1%; Score 8; DB 12; Length 8;  
Best Local Similarity 100.0%; Pred. No. 6.2e+05;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 26 VAPPEPEV 33  
Db 1 VAPPEPEV 8

RESULT 34  
US-09-833-203-40  
;; Sequence 40, Application US/09833203  
;; Publication No. US20030166277A1  
;; GENERAL INFORMATION:  
;; APPLICANT: Zauderer, Maurice  
;; APPLICANT: Smith, Ernest S.  
;; TITLE OF INVENTION: Targeted Vaccine Delivery Systems  
;; FILE REFERENCE: 1821.0020001  
;; CURRENT APPLICATION NUMBER: US/09/833,203  
;; PRIOR FILING DATE: 2001-04-12  
;; PRIOR APPLICATION NUMBER: US 60/196,472  
;; PRIOR FILING DATE: 2000-04-12  
;; NUMBER OF SEQ ID NOS: 63  
;; SOFTWARE: PatentIn version 3.0  
;; SEQ ID NO 40  
;; LENGTH: 8  
;; TYPE: PRT  
;; ORGANISM: Artificial Sequence  
;; FEATURE:  
;; NAME/KEY: misc feature  
;; OTHER INFORMATION: C35 peptides  
US-09-833-203-40

Query Match 6.1%; Score 8; DB 12; Length 8;  
Best Local Similarity 100.0%; Pred. No. 6.2e+05;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 32 EVEPGSGV 39  
Db 1 EVEPGSGV 8

```
RESULT 35
US-09-833-203-42
; Sequence 42, Application US/09833203
; Publication No. US20030166277A1
; GENERAL INFORMATION:
; APPLICANT: Zauderer, Maurice
; TITLE OF INVENTION: Targeted Vaccine Delivery Systems
; FILE REFERENCE: 1821.0020001
; CURRENT APPLICATION NUMBER: US/09/833,203
; CURRENT FILING DATE: 2001-04-12
; PRIOR APPLICATION NUMBER: US 60/196,472
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 42
; LENGTH: 8
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: C35 peptides
US-09-833-203-42

Query Match
Best Local Similarity 100.0%; Score 8; DB 12; Length 8;
Pred. No. 6.2e+05;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 52 EATYLELA 59
Db 1 EATYLELA 8

RESULT 36
US-09-833-203-45
; Sequence 45, Application US/09833203
; Publication No. US20030166277A1
; GENERAL INFORMATION:
; APPLICANT: Zauderer, Maurice
; TITLE OF INVENTION: Targeted Vaccine Delivery Systems
; FILE REFERENCE: 1821.0020001
; CURRENT APPLICATION NUMBER: US/09/833,203
; CURRENT FILING DATE: 2001-04-12
; PRIOR APPLICATION NUMBER: US 60/196,472
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 45
; LENGTH: 8
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: C35 peptides
US-09-833-203-45

Query Match
Best Local Similarity 100.0%; Score 8; DB 12; Length 8;
Pred. No. 6.2e+05;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 55 YLELASAV 62
Db 1 YLELASAV 8

RESULT 37
US-09-833-203-48
; Sequence 48, Application US/09833203
; Publication No. US20030166277A1
; GENERAL INFORMATION:
```

```
; APPLICANT: Zauderer, Maurice
; TITLE OF INVENTION: Targeted Vaccine Delivery Systems
; FILE REFERENCE: 1821.0020001
; CURRENT APPLICATION NUMBER: US/09/833,203
; CURRENT FILING DATE: 2001-04-12
; PRIOR APPLICATION NUMBER: US 60/196,472
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 48
; LENGTH: 8
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: C35 peptides
US-09-833-203-48

Query Match
Best Local Similarity 100.0%; Score 8; DB 12; Length 8;
Pred. No. 6.2e+05;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 68 GIEIESRL 75
Db 1 GIEIESRL 8

RESULT 38
US-09-833-203-52
; Sequence 52, Application US/09833203
; Publication No. US20030166277A1
; GENERAL INFORMATION:
; APPLICANT: Zauderer, Maurice
; TITLE OF INVENTION: Targeted Vaccine Delivery Systems
; FILE REFERENCE: 1821.0020001
; CURRENT APPLICATION NUMBER: US/09/833,203
; CURRENT FILING DATE: 2001-04-12
; PRIOR APPLICATION NUMBER: US 60/196,472
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 52
; LENGTH: 8
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: C35 peptides
US-09-833-203-52

Query Match
Best Local Similarity 100.0%; Score 8; DB 12; Length 8;
Pred. No. 6.2e+05;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 82 EIEINGQL 89
Db 1 EIEINGQL 8

RESULT 39
US-09-833-203-55
; Sequence 55, Application US/09833203
; Publication No. US20030166277A1
; GENERAL INFORMATION:
; APPLICANT: Zauderer, Maurice
; TITLE OF INVENTION: Targeted Vaccine Delivery Systems
; FILE REFERENCE: 1821.0020001
; CURRENT APPLICATION NUMBER: US/09/833,203
; CURRENT FILING DATE: 2001-04-12
; PRIOR APPLICATION NUMBER: US 60/196,472
```

PRIOR FILING DATE: 2000-04-12  
NUMBER OF SEQ ID NOS: 63  
SOFTWARE: Patentin version 3.0  
SEQ ID NO 55  
LENGTH: 8  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
NAME/KEY: misc\_feature  
OTHER INFORMATION: C35 peptides  
US-09-833-203-55

Query Match  
Best Local Similarity 100.0%; Pred. No. 6.2e+05;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 105 LIBAIRRA 112  
Db 1 LIBAIRRA 8

RESULT 40  
US-09-833-203-57  
Sequence 57, Application US/09833203  
Publication No. US20030166277A1  
GENERAL INFORMATION:  
APPLICANT: Zauderer, Maurice  
APPLICANT: Smith, Ernest S.  
TITLE OF INVENTION: Targeted Vaccine Delivery Systems  
FILE REFERENCE: 1821.0020001  
CURRENT APPLICATION NUMBER: US/09/833,203  
CURRENT FILING DATE: 2001-04-12  
PRIOR APPLICATION NUMBER: US 60/196,472  
PRIOR FILING DATE: 2000-04-12  
NUMBER OF SEQ ID NOS: 63  
SOFTWARE: Patentin version 3.0  
SEQ ID NO 57  
LENGTH: 8  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
NAME/KEY: misc\_feature  
OTHER INFORMATION: C35 peptides  
US-09-833-203-57

Query Match  
Best Local Similarity 100.0%; Pred. No. 6.2e+05;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 111 RASNGEYL 118  
Db 1 RASNGEYL 8

RESULT 41  
US-09-824-787B-94  
Sequence 94, Application US/09824787B  
Patent No. US20020155447A1  
GENERAL INFORMATION:  
APPLICANT: Zauderer, Maurice  
APPLICANT: Evans, Elizabeth E.  
APPLICANT: Borrello, Melinda A.  
TITLE OF INVENTION: A Gene Differentially Expressed in Breast and  
FILE REFERENCE: 1821.0040001  
CURRENT APPLICATION NUMBER: US/09/824,787B  
CURRENT FILING DATE: 2001-04-04  
PRIOR APPLICATION NUMBER: 60/194,463  
PRIOR FILING DATE: 2000-04-04  
NUMBER OF SEQ ID NOS: 147  
SOFTWARE: Patentin Ver. 2.1  
SEQ ID NO 94  
LENGTH: 9

TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-824-787B-94

Query Match  
Best Local Similarity 100.0%; Pred. No. 6.2e+05;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 65 QYPGIEIE 72  
Db 1 QYPGIEIE 8

RESULT 42  
US-09-824-787B-96  
Sequence 96, Application US/09824787B  
Patent No. US20020155447A1  
GENERAL INFORMATION:  
APPLICANT: Zauderer, Maurice  
APPLICANT: Evans, Elizabeth E.  
APPLICANT: Borrello, Melinda A.  
TITLE OF INVENTION: A Gene Differentially Expressed in Breast and  
FILE REFERENCE: 1821.0040001  
CURRENT APPLICATION NUMBER: US/09/824,787B  
CURRENT FILING DATE: 2001-04-04  
PRIOR APPLICATION NUMBER: 60/194,463  
PRIOR FILING DATE: 2000-04-04  
NUMBER OF SEQ ID NOS: 147  
SOFTWARE: Patentin Ver. 2.1  
SEQ ID NO 96  
LENGTH: 9  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-824-787B-96

Query Match  
Best Local Similarity 100.0%; Pred. No. 6.2e+05;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 93 KLENGGFP 100  
Db 1 KLENGGFP 8

RESULT 43  
US-09-824-787B-104  
Sequence 104, Application US/09824787B  
Patent No. US20020155447A1  
GENERAL INFORMATION:  
APPLICANT: Zauderer, Maurice  
APPLICANT: Evans, Elizabeth E.  
APPLICANT: Borrello, Melinda A.  
TITLE OF INVENTION: A Gene Differentially Expressed in Breast and  
FILE REFERENCE: 1821.0040001  
CURRENT APPLICATION NUMBER: US/09/824,787B  
CURRENT FILING DATE: 2001-04-04  
PRIOR APPLICATION NUMBER: 60/194,463  
PRIOR FILING DATE: 2000-04-04  
NUMBER OF SEQ ID NOS: 147  
SOFTWARE: Patentin Ver. 2.1  
SEQ ID NO 104  
LENGTH: 9  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-824-787B-104

Query Match  
Best Local Similarity 100.0%; Pred. No. 6.2e+05;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 73 SRLGGTGA 80

```
Db      1  |||||
          1 SRGCTGA 8

RESULT 44
US-09-824-787B-106
; Sequence 106, Application US/09824787B
; Patent No. US20020155447A1
; GENERAL INFORMATION:
; APPLICANT: Zauderer, Maurice
; APPLICANT: Evans, Elizabeth E.
; APPLICANT: Borrello, Melinda A.
; TITLE OF INVENTION: A Gene Differentially Expressed in Breast and
; TITLE OF INVENTION: Bladder Cancer, and Encoded Polypeptides
; FILE REFERENCE: 1821.0040001
; CURRENT APPLICATION NUMBER: US/09/824,787B
; CURRENT FILING DATE: 2001-04-04
; PRIOR APPLICATION NUMBER: 60/194,463
; PRIOR FILING DATE: 2000-04-04
; NUMBER OF SEQ ID NOS: 147
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 106
; LENGTH: 9
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-824-787B-106

Query Match
Best Local Similarity 100.0%; Score 8; DB 10; Length 9;
Pred. No. 6.2e+05;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      31  |||||
          1 EEVEPGSG 8

RESULT 45
US-09-824-787B-108
; Sequence 108, Application US/09824787B
; Patent No. US20020155447A1
; GENERAL INFORMATION:
; APPLICANT: Zauderer, Maurice
; APPLICANT: Evans, Elizabeth E.
; APPLICANT: Borrello, Melinda A.
; TITLE OF INVENTION: A Gene Differentially Expressed in Breast and
; TITLE OF INVENTION: Bladder Cancer, and Encoded Polypeptides
; FILE REFERENCE: 1821.0040001
; CURRENT APPLICATION NUMBER: US/09/824,787B
; CURRENT FILING DATE: 2001-04-04
; PRIOR APPLICATION NUMBER: 60/194,463
; PRIOR FILING DATE: 2000-04-04
; NUMBER OF SEQ ID NOS: 147
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 108
; LENGTH: 9
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-824-787B-108

Query Match
Best Local Similarity 100.0%; Score 8; DB 10; Length 9;
Pred. No. 6.2e+05;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      33  |||||
          1 VEPGSGVR 8

RESULT 46
US-09-824-787B-110
; Sequence 110, Application US/09824787B
; Patent No. US20020155447A1
; GENERAL INFORMATION:
```

```
; APPLICANT: Zauderer, Maurice
; APPLICANT: Evans, Elizabeth E.
; APPLICANT: Borrello, Melinda A.
; TITLE OF INVENTION: A Gene Differentially Expressed in Breast and
; TITLE OF INVENTION: Bladder Cancer, and Encoded Polypeptides
; FILE REFERENCE: 1821.0040001
; CURRENT APPLICATION NUMBER: US/09/824,787B
; CURRENT FILING DATE: 2001-04-04
; PRIOR APPLICATION NUMBER: 60/194,463
; PRIOR FILING DATE: 2000-04-04
; NUMBER OF SEQ ID NOS: 147
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 110
; LENGTH: 9
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-824-787B-110

Query Match
Best Local Similarity 100.0%; Score 8; DB 10; Length 9;
Pred. No. 6.2e+05;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      51  |||||
          1 FEATYDEL 8

RESULT 47
US-09-824-787B-117
; Sequence 117, Application US/09824787B
; Patent No. US20020155447A1
; GENERAL INFORMATION:
; APPLICANT: Zauderer, Maurice
; APPLICANT: Evans, Elizabeth E.
; APPLICANT: Borrello, Melinda A.
; TITLE OF INVENTION: A Gene Differentially Expressed in Breast and
; TITLE OF INVENTION: Bladder Cancer, and Encoded Polypeptides
; FILE REFERENCE: 1821.0040001
; CURRENT APPLICATION NUMBER: US/09/824,787B
; CURRENT FILING DATE: 2001-04-04
; PRIOR APPLICATION NUMBER: 60/194,463
; PRIOR FILING DATE: 2000-04-04
; NUMBER OF SEQ ID NOS: 147
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 117
; LENGTH: 9
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-824-787B-117

Query Match
Best Local Similarity 100.0%; Score 8; DB 10; Length 9;
Pred. No. 6.2e+05;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      116  |||||
          1 ETELEKTN 8

RESULT 48
US-09-824-787B-120
; Sequence 120, Application US/09824787B
; Patent No. US20020155447A1
; GENERAL INFORMATION:
; APPLICANT: Zauderer, Maurice
; APPLICANT: Evans, Elizabeth E.
; APPLICANT: Borrello, Melinda A.
; TITLE OF INVENTION: A Gene Differentially Expressed in Breast and
; TITLE OF INVENTION: Bladder Cancer, and Encoded Polypeptides
; FILE REFERENCE: 1821.0040001
; CURRENT APPLICATION NUMBER: US/09/824,787B
; CURRENT FILING DATE: 2001-04-04
; PRIOR APPLICATION NUMBER: 60/194,463
```

PRIOR FILING DATE: 2000-04-04  
NUMBER OF SEQ ID NOS: 147  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 120  
LENGTH: 9  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-824-787B-120

Query Match 6.1%; Score 8; DB 10; Length 9;  
Best Local Similarity 100.0%; Pred. No. 6.2e+05;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 110 RRASNGET 117  
DB 1 RRASNGET 8

RESULT 49  
US-09-824-787B-122  
Sequence 122, Application US/09824787B  
Patent No. US20020155447A1  
GENERAL INFORMATION:  
APPLICANT: Zauderer, Maurice  
APPLICANT: Evans, Elizabeth E.  
APPLICANT: Borrello, Melinda A.  
TITLE OF INVENTION: A Gene Differentially Expressed in Breast and  
FILE REFERENCE: 1821.0040001  
CURRENT APPLICATION NUMBER: US/09/824,787B  
CURRENT FILING DATE: 2001-04-04  
PRIOR APPLICATION NUMBER: 60/194,463  
PRIOR FILING DATE: 2000-04-04  
NUMBER OF SEQ ID NOS: 147  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 122  
LENGTH: 9  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-824-787B-122

Query Match 6.1%; Score 8; DB 10; Length 9;  
Best Local Similarity 100.0%; Pred. No. 6.2e+05;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 109 IRRASNGE 116  
DB 1 IRRASNGE 8

RESULT 50  
US-09-824-787B-133  
Sequence 133, Application US/09824787B  
Patent No. US20020155447A1  
GENERAL INFORMATION:  
APPLICANT: Zauderer, Maurice  
APPLICANT: Evans, Elizabeth E.  
APPLICANT: Borrello, Melinda A.  
TITLE OF INVENTION: A Gene Differentially Expressed in Breast and  
FILE REFERENCE: 1821.0040001  
CURRENT APPLICATION NUMBER: US/09/824,787B  
CURRENT FILING DATE: 2001-04-04  
PRIOR APPLICATION NUMBER: 60/194,463  
PRIOR FILING DATE: 2000-04-04  
NUMBER OF SEQ ID NOS: 147  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 133  
LENGTH: 9  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-824-787B-133

Query Match 6.1%; Score 8; DB 10; Length 9;  
Best Local Similarity 100.0%; Pred. No. 6.2e+05;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 83 IIRNGQLV 90  
DB 1 IIRNGQLV 8

Search completed: December 15, 2003, 17:12:19  
Job time : 32 secs

**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 15, 2003, 17:04:41 ; Search time 21 Seconds

(without alignments)  
263.939 Million cell updates/sec

Title: US-09-925-301-966

Perfect score: 131

Sequence: 1 AEVHTRKQGEAEPAAMSGE.....ASNGETLEKITSRPPCVII 131

Scoring table:  
OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 328717 seqs, 42310858 residues

Word size : 0

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: listing first 100 summaries

Database :

Issued Patents AA: \*  
1: /cgn2\_6/prodata/1/iaa/5A\_COMB.pep: \*  
2: /cgn2\_6/prodata/1/iaa/5B\_COMB.pep: \*  
3: /cgn2\_6/prodata/1/iaa/6A\_COMB.pep: \*  
4: /cgn2\_6/prodata/1/iaa/6B\_COMB.pep: \*  
5: /cgn2\_6/prodata/1/iaa/PCtUS\_COMB.pep: \*  
6: /cgn2\_6/prodata/1/iaa/backfile1.pep: \*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	8	6.1	168	4	US-08-311-731A-255
2	7	5.3	214	4	US-09-328-352-6695
3	7	5.3	222	4	US-09-252-991A-21848
4	7	5.3	628	4	US-09-267-311-4
5	7	5.3	705	4	US-08-311-731A-4
6	7	5.3	749	4	US-09-267-737-100
7	7	5.3	924	4	US-09-267-311-2
8	7	5.3	1729	4	US-09-252-690-2
9	7	5.3	2353	4	US-08-984-709A-50
10	6	4.6	36	4	US-09-252-063-9
11	6	4.6	51	1	US-08-188-228-32
12	6	4.6	51	1	US-08-332-643-32
13	6	4.6	51	1	US-08-332-638-32
14	6	4.6	52	4	US-09-252-063-1
15	6	4.6	52	4	US-09-252-063-2
16	6	4.6	52	4	US-09-252-063-3
17	6	4.6	52	4	US-09-252-063-4
18	6	4.6	52	4	US-09-252-063-5
19	6	4.6	52	4	US-09-252-063-6
20	6	4.6	52	4	US-09-549-872B-14
21	6	4.6	52	4	US-09-549-872B-15
22	6	4.6	69	4	US-08-311-731A-340
23	6	4.6	70	3	US-08-851-843A-190
24	6	4.6	70	3	US-08-974-549A-309
25	6	4.6	70	3	US-08-854-050-190
26	6	4.6	70	3	US-09-430-323-190
27	6	4.6	76	1	US-08-848-252-4

28	6	4.6	76	3	US-09-083-521-5	Sequence 5, App1
29	6	4.6	86	4	US-09-375-140-9	Sequence 9, App1
30	6	4.6	91	6	5284931-6	Patent No. 5284931
31	6	4.6	112	4	US-09-732-210-251	Sequence 251, App
32	6	4.6	120	3	US-08-341-560B-2	Sequence 2, App1
33	6	4.6	120	3	US-08-353-940-2	Sequence 2, App1
34	6	4.6	130	5	PCT-US93-03895-2	Sequence 2, App1
35	6	4.6	133	4	US-09-252-991A-21490	Sequence 21490, A
36	6	4.6	134	1	US-08-246-242-5	Sequence 5, App1
37	6	4.6	136	4	US-09-732-210-814	Sequence 814, App
38	6	4.6	136	4	US-09-732-210-815	Sequence 815, App
39	6	4.6	136	4	US-09-732-210-819	Sequence 819, App
40	6	4.6	159	4	US-09-252-991A-18982	Sequence 18982, A
41	6	4.6	166	4	US-09-252-991A-18518	Sequence 18518, A
42	6	4.6	192	4	US-09-252-991A-32773	Sequence 32773, A
43	6	4.6	192	4	US-09-252-991A-18164	Sequence 18164, A
44	6	4.6	197	4	US-09-107-532A-5858	Sequence 5858, Ap
45	6	4.6	199	4	US-09-107-532A-27431	Sequence 27431, A
46	6	4.6	209	4	US-09-252-991A-26201	Sequence 26201, A
47	6	4.6	210	4	US-09-107-532A-4087	Sequence 4087, Ap
48	6	4.6	213	4	US-09-107-532A-5566	Sequence 5566, Ap
49	6	4.6	230	4	US-09-252-991A-24101	Sequence 24101, A
50	6	4.6	248	1	US-08-744-026-1	Sequence 1, App1
51	6	4.6	248	2	US-09-102-732-1	Sequence 1, App1
52	6	4.6	248	3	US-08-341-560B-8	Sequence 8, App1
53	6	4.6	248	3	US-09-261-767-1	Sequence 1, App1
54	6	4.6	248	3	US-09-261-767-1	Sequence 1, App1
55	6	4.6	248	3	US-08-353-940-8	Sequence 8, App1
56	6	4.6	248	3	PCT-US93-03895-8	Sequence 8, App1
57	6	4.6	250	3	US-08-341-560B-6	Sequence 6, App1
58	6	4.6	250	3	US-08-353-940-6	Sequence 6, App1
59	6	4.6	254	5	PCT-US93-03895-6	Sequence 6, App1
60	6	4.6	254	4	US-09-328-352-7173	Sequence 7173, Ap
61	6	4.6	258	1	US-08-744-026-3	Sequence 4066, Ap
62	6	4.6	258	2	US-09-102-732-3	Sequence 3, App1
63	6	4.6	262	3	US-09-261-767-3	Sequence 3, App1
64	6	4.6	262	1	US-08-744-026-4	Sequence 4, App1
65	6	4.6	262	2	US-08-790-137-1	Sequence 1, App1
66	6	4.6	262	2	US-08-790-137-3	Sequence 3, App1
67	6	4.6	262	2	US-08-681-151-4	Sequence 4, App1
68	6	4.6	262	2	US-08-681-151-4	Sequence 4, App1
69	6	4.6	262	2	US-09-102-732-4	Sequence 4, App1
70	6	4.6	262	2	US-08-824-874-4	Sequence 4, App1
71	6	4.6	262	3	US-08-807-151-4	Sequence 4, App1
72	6	4.6	262	3	US-09-261-767-4	Sequence 4, App1
73	6	4.6	262	3	US-09-210-084-4	Sequence 4, App1
74	6	4.6	262	4	US-09-478-957-4	Sequence 4, App1
75	6	4.6	268	4	US-09-764-762-4	Sequence 4, App1
76	6	4.6	271	4	US-09-461-325-346	Sequence 346, App
77	6	4.6	286	5	US-09-107-532A-3966	Sequence 3966, Ap
78	6	4.6	287	4	PCT-US92-00282-9	Sequence 9, App1
79	6	4.6	289	4	US-09-305-856B-10	Sequence 10, App1
80	6	4.6	289	4	US-09-254-504-4	Sequence 4, App1
81	6	4.6	327	4	US-09-107-532A-5843	Sequence 8, App1
82	6	4.6	329	4	US-09-180-827-8	Sequence 8, App1
83	6	4.6	329	4	US-09-180-827-11	Sequence 11, App1
84	6	4.6	329	4	US-09-180-827-11	Sequence 11, App1
85	6	4.6	330	4	US-09-301-666A-10	Sequence 10, App1
86	6	4.6	330	4	US-09-252-991A-21326	Sequence 21326, A
87	6	4.6	342	4	US-09-134-001C-4190	Sequence 4190, Ap
88	6	4.6	343	4	US-09-801-861-2	Sequence 2, App1
89	6	4.6	349	4	US-09-107-532A-9914	Sequence 9914, Ap
90	6	4.6	357	2	US-08-978-182-4	Sequence 3914, Ap
91	6	4.6	357	2	US-09-205-681-4	Sequence 4, App1
92	6	4.6	357	4	US-09-252-991A-20860	Sequence 20860, A
93	6	4.6	365	4	US-09-801-874-5	Sequence 5, App1
94	6	4.6	367	4	US-09-134-001C-4168	Sequence 4168, Ap
95	6	4.6	368	4	US-09-107-532A-5250	Sequence 5250, Ap
96	6	4.6	374	4	US-09-342-681C-4	Sequence 4, App1
97	6	4.6	377	4	US-09-342-681C-4	Sequence 4, App1
98	6	4.6	388	4	US-09-252-991A-30472	Sequence 30472, A
99	6	4.6	388	4	US-09-252-991A-27381	Sequence 27381, A
100	6	4.6	390	4	US-09-252-991A-26543	Sequence 26543, A

## ALIGNMENTS

RESULT 1  
US-08-311-731A-255

Sequence 255, Application US/08311731A  
Patent No. 6583266  
GENERAL INFORMATION:  
APPLICANT: SMITH, DOUGLAS  
APPLICANT: MAO, JEN-I  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES  
RELATING TO MYCOBACTERIUM TUBERCULOSIS AND LAPRAE FOR  
NUMBER OF SEQUENCES: 411  
TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: WOLF, GREENFIELD & SACKS, P.C.  
STREET: 600 ATLANTIC AVENUE  
CITY: BOSTON  
STATE: MASSACHUSETTS  
COUNTRY: USA  
ZIP: 02210  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/311,731A  
FILING DATE:  
CLASSIFICATION: 530  
ATTORNEY/AGENT INFORMATION:  
NAME: GATES, EDWARD R.  
REGISTRATION NUMBER: 31,616  
REFERENCE/DOCKET NUMBER: C0044/7125  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617/720-3500  
TELEFAX: 617/720-2441  
INFORMATION FOR SEQ ID NO: 255:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 168 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: YES  
ORIGINAL SOURCE:  
ORGANISM: Mycobacterium leprae  
US-08-311-731A-255

Query Match 6.1%; Score 8; DB 4; Length 168;  
Best Local Similarity 100.0%; Pred. No. 1.5;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 36 GSGVRIV 43  
DB 39 GSGVRIV 46

RESULT 2  
US-09-328-352-6695

Sequence 6695, Application US/09328352  
Patent No. 6562958  
GENERAL INFORMATION:  
APPLICANT: Gary L. Bretton et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER  
TITLE OF INVENTION: BAUMANNI FOR DIAGNOSTICS AND THERAPEUTICS  
FILE REFERENCE: GTC99-03PA  
CURRENT APPLICATION NUMBER: US/09/328,352  
CURRENT FILING DATE: 1999-06-04  
NUMBER OF SEQ ID NOS: 8252  
SEQ ID NO 6695  
LENGTH: 214

TYPE: PRT  
ORGANISM: Acinetobacter baumannii  
US-09-328-352-6695

Query Match 5.3%; Score 7; DB 4; Length 214;  
Best Local Similarity 100.0%; Pred. No. 21;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 83 IEINGQL 89  
DB 104 IEINGQL 110

RESULT 3  
US-09-252-991A-21848

Sequence 21848, Application US/09252991A  
Patent No. 6551795  
GENERAL INFORMATION:  
APPLICANT: Marc J. Rubenfield et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
FILE REFERENCE: 107196.136  
CURRENT APPLICATION NUMBER: US/09/252,991A  
CURRENT FILING DATE: 1999-02-18  
PRIOR APPLICATION NUMBER: US 60/074,788  
PRIOR FILING DATE: 1998-02-18  
PRIOR APPLICATION NUMBER: US 60/094,190  
PRIOR FILING DATE: 1998-07-27  
NUMBER OF SEQ ID NOS: 33142  
SEQ ID NO 21848  
LENGTH: 222  
TYPE: PRT  
ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-21848

Query Match 5.3%; Score 7; DB 4; Length 222;  
Best Local Similarity 100.0%; Pred. No. 21;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 23 QTSVAP 29  
DB 157 QTSVAP 163

RESULT 4  
US-09-267-311-4

Sequence 4, Application US/09267311  
Patent No. 6440715  
GENERAL INFORMATION:  
APPLICANT: XU, Shuang-Yong  
TITLE OF INVENTION: Method for Cloning and Expression Of Rhodothermus  
TITLE OF INVENTION: Obamensis DNA Polymerase I Large Fragment In E. Coli  
FILE REFERENCE: NEB-157  
CURRENT APPLICATION NUMBER: US/09/267,311  
CURRENT FILING DATE: 1999-03-12  
NUMBER OF SEQ ID NOS: 15  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 4  
LENGTH: 628  
TYPE: PRT  
ORGANISM: Rhodothermus obamensis  
US-09-267-311-4

Query Match 5.3%; Score 7; DB 4; Length 628;  
Best Local Similarity 100.0%; Pred. No. 58;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 29 PPEVERP 35  
DB 588 PPEVERP 594

RESULT 5

US-08-311-731A-4  
; Sequence 4, Application US/08311731A  
; Patent No. 6583266  
; GENERAL INFORMATION:  
; APPLICANT: SMITH, DOUGLAS  
; APPLICANT: MAO, JEN-I  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES  
; TITLE OF INVENTION: RELATING TO MYCOBACTERIUM TUBERCULOSIS AND LAPRAE FOR  
; NUMBER OF SEQUENCES: 411  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: WOLF, GREENFIELD & SACKS, P.C.  
; STREET: 600 ATLANTIC AVENUE  
; CITY: BOSTON  
; STATE: MASSACHUSETTS  
; COUNTRY: USA  
; ZIP: 02210  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/311,731A  
; FILING DATE:  
; CLASSIFICATION: 530  
; ATTORNEY/AGENT INFORMATION:  
; NAME: GATES, EDWARD R.  
; REGISTRATION NUMBER: 31,616  
; REFERENCE/DOCKET NUMBER: C0044/7125  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 617/720-3500  
; TELEFAX: 617/720-2441  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 705 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; HYPOTHEICAL: YES  
; ORIGINAL SOURCE:  
; ORGANISM: MYCOBACTERIUM TUBERCULOSIS  
; US-08-311-731A-4  
Query Match 5.3%; Score 7; DB 4; Length 705;  
Best Local Similarity 100.0%; Pred. No. 65;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 34 EPGSGVR 40  
Db 502 EPGSGVR 508  
RESULT 6  
US-09-562-737-100  
; Sequence 100, Application US/09562737  
; Patent No. 6428967  
; GENERAL INFORMATION:  
; APPLICANT: Herz, Joachim  
; APPLICANT: Gotthardt, Michael  
; TITLE OF INVENTION: LDL Receptor Signaling Pathways  
; FILE REFERENCE: UTSW0708  
; CURRENT APPLICATION NUMBER: US/09/562,737  
; CURRENT FILING DATE: 2000-05-01  
; NUMBER OF SEQ ID NOS: 132  
; SOFTWARE: Patentin Ver. 2.1  
; SEQ ID NO 100  
; LENGTH: 749  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic  
; OTHER INFORMATION: Sequence

US-09-562-737-100  
Query Match 5.3%; Score 7; DB 4; Length 749;  
Best Local Similarity 100.0%; Pred. No. 68;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 82 EIEINQ 88  
Db 700 EIEINQ 706  
RESULT 7  
US-09-267-311-2  
; Sequence 2, Application US/09267311  
; Patent No. 6440715  
; GENERAL INFORMATION:  
; APPLICANT: XU, Shuang-yong  
; TITLE OF INVENTION: Method for Cloning And Expression Of Rhodothermus  
; TITLE OF INVENTION: Obamensis DNA Polymerase I Large Fragment In E. Coli  
; FILE REFERENCE: NEB-157  
; CURRENT APPLICATION NUMBER: US/09/267,311  
; CURRENT FILING DATE: 1999-03-12  
; NUMBER OF SEQ ID NOS: 15  
; SOFTWARE: Patentin Ver. 2.0  
; SEQ ID NO 2  
; LENGTH: 924  
; TYPE: PRT  
; ORGANISM: Rhodothermus obamensis  
; US-09-267-311-2  
Query Match 5.3%; Score 7; DB 4; Length 924;  
Best Local Similarity 100.0%; Pred. No. 84;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 29 PPEVEP 35  
Db 884 PPEVEP 890  
RESULT 8  
US-09-553-690-2  
; Sequence 2, Application US/09553690  
; Patent No. 6476296  
; GENERAL INFORMATION:  
; APPLICANT: Fischer, Robert L.  
; APPLICANT: Choi, Yoonhee  
; APPLICANT: Hannon, Mike  
; TITLE OF INVENTION: The Regents of the University of California  
; TITLE OF INVENTION: Nucleic Acids That Control Seed and  
; FILE REFERENCE: 023070-099900US  
; CURRENT APPLICATION NUMBER: US/09/553,690  
; CURRENT FILING DATE: 2000-04-21  
; NUMBER OF SEQ ID NOS: 50  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 2  
; LENGTH: 1729  
; TYPE: PRT  
; ORGANISM: Arabidopsis sp.  
; FEATURE:  
; OTHER INFORMATION: ATRPOS (ATR) amino acid sequence  
; US-09-553-690-2  
Query Match 5.3%; Score 7; DB 4; Length 1729;  
Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 107 EAIRAS 113  
Db 1219 EAIRAS 1225  
RESULT 9

US-08-984-709A-50  
; Sequence 50, Application US/08984709A  
; Patent No. 6320032  
; GENERAL INFORMATION:  
; APPLICANT: Williams, Mark E.  
; APPLICANT: Stauderman, Kenneth A.  
; APPLICANT: Harpold, Michael M.  
; TITLE OF INVENTION: HUMAN CALCIUM CHANNEL COMPOSITIONS AND  
; TITLE OF INVENTION: METHODS  
; NUMBER OF SEQUENCES: 52  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Heller Ehrman White & McNuliffe  
; STREET: 4250 Executive Square, Suite 700  
; City: La Jolla  
; STATE: California  
; COUNTRY: US  
; ZIP: 92037  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FASTSEQ Version 1.5  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/984,709A  
; FILING DATE: 02-DEC-1997  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Seidman, Stephanie L.  
; REGISTRATION NUMBER: 33,779  
; REFERENCE/DOCKET NUMBER: 24735-9815 (formerly 6362-9815)  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (619) 450-8400  
; TELEFAX: (619) 587-5360  
; INFORMATION FOR SEQ ID NO: 50:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2353 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; HYPOTHEICAL: NO  
; ANTI-SENSE: NO  
; FRAGMENT TYPE: internal  
; ORIGINAL SOURCE:  
; US-08-984-709A-50

Query Match 5.3%; Score 7; DB 4; Length 2353;  
Best Local Similarity 100.0%; Pred. No. 2e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 GPEAEP 15  
Db 2172 GPEAEP 2178

RESULT 10  
; US-09-252-063-9  
; Sequence 9, Application US/09252063  
; Patent No. 6538022  
; GENERAL INFORMATION:  
; APPLICANT: Pollesello, Piero  
; APPLICANT: Ovaekka, Martti  
; APPLICANT: Tenhunen, Jukka  
; APPLICANT: Viidgren, Jukka  
; APPLICANT: Yliperttula-Ikonen, Marjo  
; APPLICANT: Tiihmann, Carola  
; APPLICANT: Lotta, Timo  
; APPLICANT: Kaitola, Juha  
; TITLE OF INVENTION: Compounds for Deactivating Phospholamban Function on  
; FILE REFERENCE: 1102.0250001  
; CURRENT APPLICATION NUMBER: US/09/252,063  
; CURRENT FILING DATE: 1999-02-18

EARLIER APPLICATION NUMBER: 08/937,117  
; EARLIER FILING DATE: 1997-09-24  
; NUMBER OF SEQ ID NOS: 10  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 9  
; LENGTH: 36  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: cystolic part  
; OTHER INFORMATION: of phospholamban peptide  
; US-09-252-063-9

Query Match 4.6%; Score 6; DB 4; Length 36;  
Best Local Similarity 100.0%; Pred. No. 40;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 108 AIRRAS 113  
Db 11 AIRRAS 16

RESULT 11  
; US-08-188-228-32  
; Sequence 32, Application US/08188228  
; Patent No. 5597725  
; GENERAL INFORMATION:  
; APPLICANT: Suzuki, Shintaro  
; TITLE OF INVENTION: CADHERIN MATERIALS AND METHODS  
; NUMBER OF SEQUENCES: 62  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray &  
; ADDRESS: Borun  
; STREET: 6300 Sears Tower, 233 S. Wacker Drive  
; CITY: Chicago  
; STATE: Illinois  
; COUNTRY: USA  
; ZIP: 60606  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/188,228  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/049,460  
; FILING DATE: 19 APR 1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/872,643  
; FILING DATE: 17 APR 1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: No. 5597725and, Greta E.  
; REGISTRATION NUMBER: 35,302  
; REFERENCE/DOCKET NUMBER: 31340  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (312) 474-6300  
; TELEFAX: (312) 474-0448  
; TELEX: 25-3856  
; INFORMATION FOR SEQ ID NO: 32:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 51 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-188-228-32

Query Match 4.6%; Score 6; DB 1; Length 51;  
Best Local Similarity 100.0%; Pred. No. 36;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 39 VRIIVE 44  
Db 38 VRIIVE 43

## RESULT 12

US-08-332-643-32  
Sequence 32, Application US/08332643  
Patent No. 5639634  
GENERAL INFORMATION:  
APPLICANT: Suzuki, Shintaro  
TITLE OF INVENTION: CADHERIN MATERIALS AND METHODS  
NUMBER OF SEQUENCES: 56  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray &  
ADDRESSEE: Bicknell  
STREET: Two First National Plaza, 20 South Clark  
CITY: Chicago  
STATE: Illinois  
COUNTRY: USA  
ZIP: 60603  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/332,643  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/07/872,643  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: No. 5639634and, Greta E.  
REGISTRATION NUMBER: 35,302  
REFERENCE/DOCKET NUMBER: 27866/30795  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (312) 346-5750  
TELEFAX: (312) 984-9740  
TELEX: 25-3856  
INFORMATION FOR SEQ ID NO: 32:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 51 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-332-643-32

Query Match 4.6%; Score 6; DB 1; Length 51;  
Best Local Similarity 100.0%; Pred. No. 56;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 39 VRIIVE 44  
Db 38 VRIIVE 43

## RESULT 13

US-08-332-638-32  
Sequence 32, Application US/08332638  
Patent No. 5646250  
GENERAL INFORMATION:  
APPLICANT: Suzuki, Shintaro  
TITLE OF INVENTION: CADHERIN MATERIALS AND METHODS  
NUMBER OF SEQUENCES: 62  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray &  
ADDRESSEE: Borun  
STREET: 6300 Sears Tower, 233 S. Wacker Drive  
CITY: Chicago  
STATE: Illinois

COUNTRY: USA  
ZIP: 60606  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/332,638  
FILING DATE: 01-NOV-1994  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/872,643  
FILING DATE: 17 APR 1992  
APPLICATION NUMBER: US/08/049,460  
FILING DATE:

ATTORNEY/AGENT INFORMATION:  
NAME: No. 5646250and, Greta E.  
REGISTRATION NUMBER: 35,302  
REFERENCE/DOCKET NUMBER: 31340  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (312) 474-6300  
TELEFAX: (312) 474-0448  
TELEX: 25-3856  
INFORMATION FOR SEQ ID NO: 32:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 51 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-332-638-32

Query Match 4.6%; Score 6; DB 1; Length 51;  
Best Local Similarity 100.0%; Pred. No. 56;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 39 VRIIVE 44  
Db 38 VRIIVE 43

## RESULT 14

US-09-252-063-1  
Sequence 1, Application US/09252063  
Patent No. 6538022  
GENERAL INFORMATION:  
APPLICANT: Pollesello, Piero  
APPLICANT: Ovasa, Martti  
APPLICANT: Tenhunen, Jukka  
APPLICANT: Viidgren, Jukka  
APPLICANT: Yliperttula-Ikonen, Marjo  
APPLICANT: Tilmann, Carola  
APPLICANT: Lotta, Timo  
APPLICANT: Kaivola, Juna  
TITLE OF INVENTION: Compounds for Deactivating Phospholamban Function on  
FILE REFERENCE: 1102.0250001  
CURRENT APPLICATION NUMBER: US/09/252,063  
CURRENT FILING DATE: 1999-02-18  
EARLIER APPLICATION NUMBER: 08/937,117  
EARLIER FILING DATE: 1997-09-24  
NUMBER OF SEQ ID NOS: 10  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 1  
LENGTH: 52  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-252-063-1

Query Match 4.6%; Score 6; DB 4; Length 52;  
Best Local Similarity 100.0%; Pred. No. 57;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 108 AIRRAS 113  
| | | | |  
Db 11 AIRRAS 16

## RESULT 15

US-09-252-063-2  
; Sequence 2, Application US/09252063  
; Patent No. 6538022  
; GENERAL INFORMATION:  
; APPLICANT: Pollesello, Piero  
; APPLICANT: Ovaska, Martti  
; APPLICANT: Tenhunen, Jukka  
; APPLICANT: Viigren, Jukka  
; APPLICANT: Yliperttula-Ikonen, Marjo  
; APPLICANT: Tilgmann, Carola  
; APPLICANT: Lotta, Timo  
; APPLICANT: Kaivola, Juha  
; TITLE OF INVENTION: Compounds for Deactivating Phospholamban Function on  
; TITLE OF INVENTION: Ca-ATPase (Phospholamban Inhibitors)  
; FILE REFERENCE: 1102.0250001  
; CURRENT APPLICATION NUMBER: US/09/252,063  
; CURRENT FILING DATE: 1999-02-18  
; EARLIER APPLICATION NUMBER: 08/937,117  
; EARLIER FILING DATE: 1997-09-24  
; NUMBER OF SEQ ID NOS: 10  
; SOFTWARE: Patentin Ver. 2.0  
; SEQ ID NO 2  
; LENGTH: 52  
; TYPE: PRT  
; ORGANISM: Sus sp.  
US-09-252-063-2

Query Match 4.6%; Score 6; DB 4; Length 52;  
Best Local Similarity 100.0%; Pred. No. 57;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 108 AIRRAS 113  
| | | | |  
Db 11 AIRRAS 16

## RESULT 16

US-09-252-063-3  
; Sequence 3, Application US/09252063  
; Patent No. 6538022  
; GENERAL INFORMATION:  
; APPLICANT: Pollesello, Piero  
; APPLICANT: Ovaska, Martti  
; APPLICANT: Tenhunen, Jukka  
; APPLICANT: Viigren, Jukka  
; APPLICANT: Yliperttula-Ikonen, Marjo  
; APPLICANT: Tilgmann, Carola  
; APPLICANT: Lotta, Timo  
; APPLICANT: Kaivola, Juha  
; TITLE OF INVENTION: Compounds for Deactivating Phospholamban Function on  
; TITLE OF INVENTION: Ca-ATPase (Phospholamban Inhibitors)  
; FILE REFERENCE: 1102.0250001  
; CURRENT APPLICATION NUMBER: US/09/252,063  
; CURRENT FILING DATE: 1999-02-18  
; EARLIER APPLICATION NUMBER: 08/937,117  
; EARLIER FILING DATE: 1997-09-24  
; NUMBER OF SEQ ID NOS: 10  
; SOFTWARE: Patentin Ver. 2.0  
; SEQ ID NO 3  
; LENGTH: 52  
; TYPE: PRT  
; ORGANISM: Canis sp.  
US-09-252-063-3

Query Match 4.6%; Score 6; DB 4; Length 52;  
Best Local Similarity 100.0%; Pred. No. 57;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 108 AIRRAS 113  
| | | | |  
Db 11 AIRRAS 16

## RESULT 17

US-09-252-063-4  
; Sequence 4, Application US/09252063  
; Patent No. 6538022  
; GENERAL INFORMATION:  
; APPLICANT: Pollesello, Piero  
; APPLICANT: Ovaska, Martti  
; APPLICANT: Tenhunen, Jukka  
; APPLICANT: Viigren, Jukka  
; APPLICANT: Yliperttula-Ikonen, Marjo  
; APPLICANT: Tilgmann, Carola  
; APPLICANT: Lotta, Timo  
; APPLICANT: Kaivola, Juha  
; TITLE OF INVENTION: Compounds for Deactivating Phospholamban Function on  
; TITLE OF INVENTION: Ca-ATPase (Phospholamban Inhibitors)  
; FILE REFERENCE: 1102.0250001  
; CURRENT APPLICATION NUMBER: US/09/252,063  
; CURRENT FILING DATE: 1999-02-18  
; EARLIER APPLICATION NUMBER: 08/937,117  
; EARLIER FILING DATE: 1997-09-24  
; NUMBER OF SEQ ID NOS: 10  
; SOFTWARE: Patentin Ver. 2.0  
; SEQ ID NO 4  
; LENGTH: 52  
; TYPE: PRT  
; ORGANISM: Oryctolagus cuniculus  
US-09-252-063-4

Query Match 4.6%; Score 6; DB 4; Length 52;  
Best Local Similarity 100.0%; Pred. No. 57;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 108 AIRRAS 113  
| | | | |  
Db 11 AIRRAS 16

## RESULT 18

US-09-252-063-5  
; Sequence 5, Application US/09252063  
; Patent No. 6538022  
; GENERAL INFORMATION:  
; APPLICANT: Pollesello, Piero  
; APPLICANT: Ovaska, Martti  
; APPLICANT: Tenhunen, Jukka  
; APPLICANT: Viigren, Jukka  
; APPLICANT: Yliperttula-Ikonen, Marjo  
; APPLICANT: Tilgmann, Carola  
; APPLICANT: Lotta, Timo  
; APPLICANT: Kaivola, Juha  
; TITLE OF INVENTION: Compounds for Deactivating Phospholamban Function on  
; TITLE OF INVENTION: Ca-ATPase (Phospholamban Inhibitors)  
; FILE REFERENCE: 1102.0250001  
; CURRENT APPLICATION NUMBER: US/09/252,063  
; CURRENT FILING DATE: 1999-02-18  
; EARLIER APPLICATION NUMBER: 08/937,117  
; EARLIER FILING DATE: 1997-09-24  
; NUMBER OF SEQ ID NOS: 10  
; SOFTWARE: Patentin Ver. 2.0  
; SEQ ID NO 5  
; LENGTH: 52  
; TYPE: PRT  
; ORGANISM: Rattus sp.  
US-09-252-063-5

Query Match 4.6%; Score 6; DB 4; Length 52;  
Best Local Similarity 100.0%; Pred. No. 57;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;



TELEPHONE: 617/720-3500  
TELEFAX: 617/720-2441  
INFORMATION FOR SEQ ID NO: 340:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 69 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHEMETICAL: YES  
ORIGINAL SOURCE: Mycobacterium leprae  
US-08-311-731A-340

Query Match 4.6%; Score 6; DB 4; Length 69;  
Best Local Similarity 100.0%; Pred. No. 74;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 37 SGVRIV 42  
DB 63 SGVRIV 68

RESULT 23  
US-08-851-843A-190  
Sequence 190, Application US/08851843A  
Patent No. 6093809  
GENERAL INFORMATION:  
APPLICANT: Cech, Thomas R.  
APPLICANT: Lingner, Joachim  
APPLICANT: Nakamura, Toru  
APPLICANT: Chapman, Karen B.  
APPLICANT: Morin, Gregg B.  
APPLICANT: Harley, Calvin  
APPLICANT: Andrews, William H.  
TITLE OF INVENTION: No. 6093809el Telomerase  
NUMBER OF SEQUENCES: 225  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Townsend and Townsend and Crew LLP  
STREET: Two Embarcadero Center, 8th Floor  
CITY: San Francisco  
STATE: California  
COUNTRY: United States of America  
ZIP: 94111  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/851,843A  
FILING DATE: 06-MAY-1997  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/846,017  
FILING DATE: 25-APR-1997  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/844,419  
FILING DATE: 18-APR-1997  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/724,643  
FILING DATE: 01-OCT-1996  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Apple, Randolph T.  
REGISTRATION NUMBER: 36,429  
REFERENCE/DOCKET NUMBER: 015389-002930US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 576-0200  
TELEFAX: (415) 576-0300  
INFORMATION FOR SEQ ID NO: 190:  
SEQUENCE CHARACTERISTICS:

LENGTH: 70 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-851-843A-190

Query Match 4.6%; Score 6; DB 3; Length 70;  
Best Local Similarity 100.0%; Pred. No. 75;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 73 SRLGCT 78  
DB 55 SRLGCT 60

RESULT 24  
US-08-974-549A-309  
Sequence 309, Application US/08974549A  
Patent No. 6166178  
GENERAL INFORMATION:  
APPLICANT: Cech, Thomas R.  
APPLICANT: Lingner, Joachim  
APPLICANT: Nakamura, Toru  
APPLICANT: Chapman, Karen B.  
APPLICANT: Morin, Gregg B.  
APPLICANT: Harley, Calvin B.  
APPLICANT: Andrews, William H.  
TITLE OF INVENTION: Human Telomerase Catalytic Subunit  
NUMBER OF SEQUENCES: 727  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Townsend and Townsend and Crew LLP  
STREET: Two Embarcadero Center, Eighth Floor  
CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94111-3834  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/974,549A  
FILING DATE: 19-NOV-1997  
CLASSIFICATION: 536  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/724,643  
FILING DATE: 01-OCT-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/844,419  
FILING DATE: 18-APR-1997  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/846,017  
FILING DATE: 25-APR-1997  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/851,843  
FILING DATE: 06-MAY-1997  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/854,050  
FILING DATE: 09-MAY-1997  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/911,312  
FILING DATE: 14-AUG-1997  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/912,951  
FILING DATE: 14-AUG-1997  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/915,503  
FILING DATE: 14-AUG-1997  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: WO PCT/US97/17618  
FILING DATE: 01-OCT-1997



PRIOR APPLICATION DATA:  
APPLICATION NUMBER: WO PCT/US97/17885  
FILING DATE: 01-OCT-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Apple, Randolph T.  
REGISTRATION NUMBER: 36,429  
REFERENCE/DOCKET NUMBER: 015389-00261005  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 576-0200  
TELEFAX: (415) 576-0300  
INFORMATION FOR SEQ ID NO: 309:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 70 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-974-549A-309

Query Match 4.6%; Score 6; DB 3; Length 70;  
Best Local Similarity 100.0%; Pred. No. 75;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 73 SRUGT 78  
Db 55 SRLGT 60

## RESULT 25

US-08-854-050-190  
Sequence 190, Application US/08854050  
Patent No. 6261836  
GENERAL INFORMATION:  
APPLICANT: Cecch, Thomas R.  
APPLICANT: Lingner, Joachim  
APPLICANT: Nakamura, Toru  
APPLICANT: Chapman, Karen B.  
APPLICANT: Morlin, Gregg B.  
APPLICANT: Harley, Calvin  
APPLICANT: Andrews, William H.  
TITLE OF INVENTION: No. 6261836el Telomerase  
NUMBER OF SEQUENCES: 225  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Townsend and Townsend and Crew LLP  
STREET: Two Embarcadero Center, 8th Floor  
CITY: San Francisco  
STATE: California  
COUNTRY: United States of America  
ZIP: 94111  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/854,050  
FILING DATE: 09-MAY-1997  
CLASSIFICATION: 536  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/851,843  
FILING DATE: 06-MAY-1997  
CLASSIFICATION: 536  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/846,017  
FILING DATE: 25-APR-1997  
CLASSIFICATION: 536  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/844,419  
FILING DATE: 18-APR-1997  
CLASSIFICATION: 536  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/724,643  
FILING DATE: 01-OCT-1996

CLASSIFICATION: 536  
ATTORNEY/AGENT INFORMATION:  
NAME: Apple, Randolph T.  
REGISTRATION NUMBER: 36,429  
REFERENCE/DOCKET NUMBER: 015389-00293005  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 576-0200  
TELEFAX: (415) 576-0300  
INFORMATION FOR SEQ ID NO: 190:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 70 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-854-050-190

Query Match 4.6%; Score 6; DB 3; Length 70;  
Best Local Similarity 100.0%; Pred. No. 75;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 73 SRUGT 78  
Db 55 SRLGT 60

## RESULT 26

US-09-430-323-190  
Sequence 190, Application US/09430323  
Patent No. 6309867  
GENERAL INFORMATION:  
APPLICANT: Cecch, Thomas R.  
APPLICANT: Lingner, Joachim  
APPLICANT: Nakamura, Toru  
APPLICANT: Chapman, Karen B.  
APPLICANT: Morlin, Gregg B.  
APPLICANT: Harley, Calvin  
APPLICANT: Andrews, William H.  
TITLE OF INVENTION: No. 6309867el Telomerase  
NUMBER OF SEQUENCES: 225  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Townsend and Townsend and Crew LLP  
STREET: Two Embarcadero Center, 8th Floor  
CITY: San Francisco  
STATE: California  
COUNTRY: United States of America  
ZIP: 94111  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/430,323  
FILING DATE: 29-OCT-1999  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/854,050  
FILING DATE: 09-MAY-1997  
APPLICATION NUMBER: US 08/851,843  
FILING DATE: 06-MAY-1997  
APPLICATION NUMBER: US 08/846,017  
FILING DATE: 25-APR-1997  
APPLICATION NUMBER: US 08/844,419  
FILING DATE: 18-APR-1997  
APPLICATION NUMBER: US 08/724,643  
FILING DATE: 01-OCT-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Apple, Randolph T.  
REGISTRATION NUMBER: 36,429  
REFERENCE/DOCKET NUMBER: 015389-00293005  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 576-0200

TELEFAX: (415) 576-0300  
INFORMATION FOR SEQ ID NO: 190:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 70 amino acids  
TYPE: amino acid  
STRANDEDNESS: <Unknown>  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
SEQUENCE DESCRIPTION: SEQ ID NO: 190:  
US-09-430-323-190

Query Match 4.6%; Score 6; DB 4; Length 70;  
Best Local Similarity 100.0%; Pred. No. 75;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 73 SRIQGT 78  
Db 55 SRIQGT 60

RESULT 27  
US-08-848-252-4  
Sequence 4, Application US/08848252  
Patent No. 5804177  
GENERAL INFORMATION:  
APPLICANT: Humphries, Keith R.  
TITLE OF INVENTION: METHOD OF USING CD24 AS A CELL MARKER  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Bereskin & Parr  
STREET: 40 King Street West  
CITY: Toronto  
STATE: Ontario  
COUNTRY: Canada  
ZIP: M5H 3Y2  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/848,252  
FILING DATE: 29-APR-1997  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/538,052  
FILING DATE:  
APPLICATION NUMBER: US 08/151,672  
FILING DATE: 15-NOV-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: McDiarmid, Shona S.  
REGISTRATION NUMBER: P-38,798  
REFERENCE/DOCKET NUMBER: 3158-028  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (416) 364-7311  
TELEFAX: (416) 361-1398  
TELEX: 06-23115  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 76 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-848-252-4

Query Match 4.6%; Score 6; DB 1; Length 76;  
Best Local Similarity 100.0%; Pred. No. 82;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 23 QTSVAP 28  
Db 28 QTSVAP 33

RESULT 28  
US-09-083-521-5  
Sequence 5, Application US/09083521  
Patent No. 6048970  
GENERAL INFORMATION:  
APPLICANT: Lai, Preeti  
APPLICANT: Guegler, Karl J.  
APPLICANT: Corley, Neil C.  
TITLE OF INVENTION: PROSTATE GROWTH-ASSOCIATED MEMBRANE PROTEINS  
NUMBER OF SEQUENCES: 7  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: INCYTE PHARMACEUTICALS, INC.  
STREET: 3174 PORTER DRIVE  
CITY: PALO ALTO  
STATE: CALIFORNIA  
COUNTRY: USA  
ZIP: 94304

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/083,521  
FILING DATE: Herewith  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: CERRONE, MICHAEL C.  
REGISTRATION NUMBER: 39,132  
REFERENCE/DOCKET NUMBER: PF-0527 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (650) 855-0555  
TELEFAX: (650) 845-4166  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 76 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
LIBRARY: Genbank  
CLONE: 1216498  
US-09-083-521-5

Query Match 4.6%; Score 6; DB 3; Length 76;  
Best Local Similarity 100.0%; Pred. No. 82;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 23 QTSVAP 28  
Db 28 QTSVAP 33

RESULT 29  
US-09-375-140-9  
Sequence 9, Application US/09375140  
Patent No. 6488540  
GENERAL INFORMATION:  
APPLICANT: Kavanagh, T.  
APPLICANT: Lao, N.  
TITLE OF INVENTION: A NOVEL PLASTID-TARGETING NUCLEIC ACID SEQUENCE, A  
TITLE OF INVENTION: NOVEL BETA-AMYLASE SEQUENCE, A STIMULUS-RESPONSIVE  
FILE REFERENCE: 9341-017  
CURRENT APPLICATION NUMBER: US/09/375,140  
CURRENT FILING DATE: 1999-08-16  
NUMBER OF SEQ ID NOS: 11  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 9  
LENGTH: 86  
TYPE: PRT  
ORGANISM: Arabidopsis thaliana

US-09-375-140-9

Query Match 4.6%; Score 6; DB 4; Length 86;  
Best Local Similarity 100.0%; Pred. No. 92;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 115 GETLER 120  
Db 63 GETLER 68

RESULT 30  
5284931-6  
; Patent No. 5284931  
; APPLICANT: SPRINGER, TIMOTHY A., ROTHLEIN, ROBERT, MARLIN,  
; STEVEN D., DUSTIN, MICHAEL L.  
; TITLE OF INVENTION: INTERCELLULAR ADHESION MOLECULES AND  
; THEIR BINDING LIGANDS  
; NUMBER OF SEQUENCES: 41  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/515,478  
; FILING DATE: 27-APR-1990  
; SEQ ID NO: 6:  
; LENGTH: 91  
5284931-6

Query Match 4.6%; Score 6; DB 6; Length 91;  
Best Local Similarity 100.0%; Pred. No. 97;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 50 GFEATY 55  
Db 48 GFEATY 53

RESULT 31  
US-09-732-210-251  
; Sequence 251, Application US/09732210  
; Patent No. 6573361  
; GENERAL INFORMATION:  
; APPLICANT: Bunkers, Greg J.  
; APPLICANT: Liang, Jihong  
; APPLICANT: Mitanck, Cindy A.  
; APPLICANT: Seale, Jeffrey W.  
; APPLICANT: Wu, Yomte S.  
; TITLE OF INVENTION: Anti-fungal Proteins and Methods for Their Use  
; FILE REFERENCE: 38-21(15036)B  
; CURRENT APPLICATION NUMBER: US/09/732,210  
; PRIOR FILING DATE: 2000-12-07  
; PRIOR APPLICATION NUMBER: US 60/169,513  
; PRIOR FILING DATE: 1999-12-07  
; PRIOR APPLICATION NUMBER: US 60/169,340  
; PRIOR FILING DATE: 1999-12-07  
; NUMBER OF SEQ ID NOS: 1753  
; SEQ ID NO 251  
; LENGTH: 112  
; TYPE: PRT  
; ORGANISM: Chlamydia trachomatis  
US-09-732-210-251

Query Match 4.6%; Score 6; DB 4; Length 112;  
Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 59 ASAYKE 64  
Db 98 ASAYKE 103

RESULT 32  
US-08-341-560B-2  
; Sequence 2, Application US/08341560B  
; Patent No. 6165745

GENERAL INFORMATION:  
; APPLICANT: Ward, E. Sally  
; APPLICANT: Kim, Jin-Kyoo  
; TITLE OF INVENTION: Recombinant Production of  
; TITLE OF INVENTION: Immunoglobulin-like Domains in Prokaryotic Cells  
; NUMBER OF SEQUENCES: 33  
; CORRESPONDENCE ADDRESSES:  
; ADDRESSEE: ARNOLD, WHITE & DURKEE  
; STREET: P.O. Box 4433  
; CITY: Houston  
; STATE: Texas  
; COUNTRY: USA  
; ZIP: 7721-4433

COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: IBM PC compatible  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/341,560B  
; FILING DATE: 17-NOV-1994  
; CLASSIFICATION: 435

PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/963,333  
; FILING DATE: 19-OCT-1992  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/873,930  
; FILING DATE: 24-APR-1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Parker, David L.  
; REGISTRATION NUMBER: 32,165  
; REFERENCE/DOCKET NUMBER: UTSD:353  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (512) 418-3000  
; TELEFAX: (713) 789-2679

INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 120 amino acids  
; TYPE: amino acid  
; STRANDEDNESS:  
; TOPOLOGY: linear  
US-08-341-560B-2

Query Match 4.6%; Score 6; DB 3; Length 120;  
Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 50 GFEATY 55  
Db 60 GFEATY 65

RESULT 33  
US-08-353-940-2  
; Sequence 2, Application US/08353940  
; Patent No. 6393368  
; GENERAL INFORMATION:  
; APPLICANT: WARD, ELIZABETH SALLY  
; TITLE OF INVENTION: SECRETION OF T CELL RECEPTOR  
; TITLE OF INVENTION: FRAGMENTS FROM RECOMBINANT  
; TITLE OF INVENTION: ESCHERICHIA COLI CELLS  
; NUMBER OF SEQUENCES: 8  
; CORRESPONDENCE ADDRESSES:  
; ADDRESSEE: ARNOLD, WHITE & DURKEE  
; STREET: P.O. BOX 4433  
; CITY: HOUSTON  
; STATE: TEXAS  
; COUNTRY: USA  
; ZIP: 77210

COMPUTER READABLE FORM:  
; MEDIUM TYPE: FLOPPY DISK  
; OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: WORDPERFECT 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/353,940  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/873,930  
FILING DATE: 04/24/92  
ATTORNEY/AGENT INFORMATION:  
NAME: KITCHELL, BARBARA S.  
REGISTRATION NUMBER: 33,928  
REFERENCE/DOCKET NUMBER: UTSD:293/KIT  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 512-474-7200  
TELEFAX: 512-474-7577  
TELEX: NOT APPLICABLE  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 120 amino acid residues  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-353-940-2

Query Match 4.6%; Score 6; DB 4; Length 120;  
Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 50 GFEATY 55  
Db 60 GFEATY 65

RESULT 34  
PCT-US93-03895-2  
Sequence 2, Application PC/TUS9303895  
GENERAL INFORMATION:  
APPLICANT: BOARD OF REGENTS, THE UNIVERSITY  
APPLICANT: OF TEXAS SYSTEM  
TITLE OF INVENTION: RECOMBINANT PRODUCTION OF  
TITLE OF INVENTION: IMMUNOGLOBULIN-LIKE DOMAINS IN  
TITLE OF INVENTION: PROKARYOTIC CELLS  
NUMBER OF SEQUENCES: 16  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: ARNOLD, WHITE & DURKEE  
STREET: P.O. BOX 4433  
CITY: HOUSTON  
STATE: TEXAS  
COUNTRY: USA  
ZIP: 77210  
COMPUTER READABLE FORM:  
MEDIUM TYPE: FLOPPY DISK  
COMPUTER: IBM PC COMPATIBLE  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: WORDPERFECT 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US93/03895  
FILING DATE: 19930426  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: KITCHELL, BARBARA S.  
REGISTRATION NUMBER: 33,928  
REFERENCE/DOCKET NUMBER: UTPD53PCT  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 512-320-7200  
TELEFAX: 512-474-7577  
TELEX: 79-0924  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 120 amino acid residues  
TYPE: AMINO ACID  
STRANDEDNESS: single

TOPOLOGY: linear  
PCT-US93-03895-2

Query Match 4.6%; Score 6; DB 5; Length 120;  
Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 50 GFEATY 55  
Db 60 GFEATY 65

RESULT 35  
US-09-252-991A-21490  
Sequence 21490, Application US/09252991A  
Patent No. 6551795  
GENERAL INFORMATION:  
APPLICANT: Marc J. Rubenfield et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
FILE REFERENCE: 107196.136  
CURRENT APPLICATION NUMBER: US/09/252,991A  
CURRENT FILING DATE: 1999-02-18  
PRIOR APPLICATION NUMBER: US 60/074,788  
PRIOR FILING DATE: 1998-02-18  
PRIOR APPLICATION NUMBER: US 60/094,190  
PRIOR FILING DATE: 1998-07-27  
NUMBER OF SEQ ID NOS: 33142  
SEQ ID NO 21490  
LENGTH: 133  
TYPE: PRT  
ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-21490

Query Match 4.6%; Score 6; DB 4; Length 133;  
Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 18 SGEPRQ 23  
Db 11 SGEPRQ 16

RESULT 36  
US-08-246-242-5  
Sequence 5, Application US/08246242  
Patent No. 5675060  
GENERAL INFORMATION:  
APPLICANT: Benoist, C.  
APPLICANT: Mathis, D.  
APPLICANT: Kouskoff, V.  
TITLE OF INVENTION: Transgenic Arthritic Mice  
NUMBER OF SEQUENCES: 9  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sterne, Kessler, Goldstein & Fox  
STREET: 1100 New York Avenue, Suite 600  
CITY: Washington  
STATE: D.C.  
COUNTRY: U.S.A.  
ZIP: 20005  
COMPUTER READABLE FORM:  
MEDIUM TYPE: FLOPPY disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/246,242  
FILING DATE: 19-MAR-1994  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Goldstein, Jorge A.  
REGISTRATION NUMBER: 29,021  
REFERENCE/DOCKET NUMBER: 1383.0080000

TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 371-2600  
TELEFAX: (202) 371-2540  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 134 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-246-242-5

Query Match  
Best Local Similarity 100.0%; Score 6; DB 1; Length 134;  
Pred. No. 1.4e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 50 GFEATY 55  
Db 83 GFEATY 88

RESULT 37  
US-09-732-210-814  
Sequence 814, Application US/09732210  
Patent No. 6573361  
GENERAL INFORMATION:  
APPLICANT: Bunkers, Greg J.  
APPLICANT: Liang, Jihong  
APPLICANT: Mitcanck, Cindy A.  
APPLICANT: Seale, Jeffrey W.  
APPLICANT: Wu, Yonnie S.  
TITLE OF INVENTION: Anti-fungal Proteins and Methods for Their Use  
FILE REFERENCE: 38-21(15036)B  
CURRENT APPLICATION NUMBER: US/09/732,210  
CURRENT FILING DATE: 2000-12-07  
PRIOR APPLICATION NUMBER: US 60/169,513  
PRIOR FILING DATE: 1999-12-07  
PRIOR APPLICATION NUMBER: US 60/169,340  
PRIOR FILING DATE: 1999-12-07  
NUMBER OF SEQ ID NOS: 1753  
SEQ ID NO 814  
LENGTH: 136  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-732-210-814

Query Match  
Best Local Similarity 100.0%; Score 6; DB 4; Length 136;  
Pred. No. 1.4e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 108 AIRRAS 113  
Db 109 AIRRAS 114

RESULT 38  
US-09-732-210-815  
Sequence 815, Application US/09732210  
Patent No. 6573361  
GENERAL INFORMATION:  
APPLICANT: Bunkers, Greg J.  
APPLICANT: Liang, Jihong  
APPLICANT: Mitcanck, Cindy A.  
APPLICANT: Seale, Jeffrey W.  
APPLICANT: Wu, Yonnie S.  
TITLE OF INVENTION: Anti-fungal Proteins and Methods for Their Use  
FILE REFERENCE: 38-21(15036)B  
CURRENT APPLICATION NUMBER: US/09/732,210  
CURRENT FILING DATE: 2000-12-07  
PRIOR APPLICATION NUMBER: US 60/169,513  
PRIOR FILING DATE: 1999-12-07  
PRIOR APPLICATION NUMBER: US 60/169,340  
PRIOR FILING DATE: 1999-12-07  
NUMBER OF SEQ ID NOS: 1753

SEQ ID NO 815  
LENGTH: 136  
TYPE: PRT  
ORGANISM: Mus musculus  
US-09-732-210-815

Query Match  
Best Local Similarity 100.0%; Score 6; DB 4; Length 136;  
Pred. No. 1.4e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 108 AIRRAS 113  
Db 109 AIRRAS 114

RESULT 39  
US-09-732-210-819  
Sequence 819, Application US/09732210  
Patent No. 6573361  
GENERAL INFORMATION:  
APPLICANT: Bunkers, Greg J.  
APPLICANT: Liang, Jihong  
APPLICANT: Mitcanck, Cindy A.  
APPLICANT: Seale, Jeffrey W.  
APPLICANT: Wu, Yonnie S.  
TITLE OF INVENTION: Anti-fungal Proteins and Methods for Their Use  
FILE REFERENCE: 38-21(15036)B  
CURRENT APPLICATION NUMBER: US/09/732,210  
CURRENT FILING DATE: 2000-12-07  
PRIOR APPLICATION NUMBER: US 60/169,513  
PRIOR FILING DATE: 1999-12-07  
PRIOR APPLICATION NUMBER: US 60/169,340  
PRIOR FILING DATE: 1999-12-07  
NUMBER OF SEQ ID NOS: 1753  
SEQ ID NO 819  
LENGTH: 136  
TYPE: PRT  
ORGANISM: Rattus norvegicus  
US-09-732-210-819

Query Match  
Best Local Similarity 100.0%; Score 6; DB 4; Length 136;  
Pred. No. 1.4e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 108 AIRRAS 113  
Db 109 AIRRAS 114

RESULT 40  
US-09-252-991A-18982  
Sequence 18982, Application US/09252991A  
Patent No. 6551795  
GENERAL INFORMATION:  
APPLICANT: Marc J. Rubenfield et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
FILE REFERENCE: 107196.136  
CURRENT APPLICATION NUMBER: US/09/252,991A  
CURRENT FILING DATE: 1999-02-18  
PRIOR APPLICATION NUMBER: US 60/074,788  
PRIOR FILING DATE: 1998-02-18  
PRIOR APPLICATION NUMBER: US 60/094,190  
PRIOR FILING DATE: 1998-07-27  
NUMBER OF SEQ ID NOS: 33142  
SEQ ID NO 18982  
LENGTH: 159  
TYPE: PRT  
ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-18982

Query Match  
Best Local Similarity 100.0%; Score 6; DB 4; Length 159;  
Pred. No. 1.7e+02;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 73 SRLGTT 78  
|||||

Db 126 SRLGTT 131

RESULT 41

US-09-252-991A-18518  
; Sequence 18518, Application US/09252991A  
; Patent No. 6551795

GENERAL INFORMATION:

APPLICANT: Marc J. Rubenfield et al.

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

CURRENT APPLICATION NUMBER: US/09/252,991A

PRIOR FILING DATE: 1999-02-18

PRIOR APPLICATION NUMBER: US 60/074,788

PRIOR FILING DATE: 1998-07-27

NUMBER OF SEQ ID NOS: 33142

SEQ ID NO 18518

LENGTH: 166

TYPE: PRT

ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-18518

Query Match

Best Local Similarity 100.0%; Pred. No. 1.7e+02;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 10 PEAPEA 15  
|||||

Db 138 PEAPEA 143

RESULT 42

US-09-252-991A-32773  
; Sequence 32773, Application US/09252991A  
; Patent No. 6551795

GENERAL INFORMATION:

APPLICANT: Marc J. Rubenfield et al.

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

CURRENT APPLICATION NUMBER: US/09/252,991A

PRIOR FILING DATE: 1999-02-18

PRIOR APPLICATION NUMBER: US 60/074,788

PRIOR FILING DATE: 1998-07-27

NUMBER OF SEQ ID NOS: 33142

SEQ ID NO 32773

LENGTH: 191

TYPE: PRT

ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-32773

Query Match

Best Local Similarity 100.0%; Pred. No. 2e+02;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

GENERAL INFORMATION:

APPLICANT: Marc J. Rubenfield et al.

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

CURRENT APPLICATION NUMBER: US/09/252,991A

PRIOR FILING DATE: 1999-02-18

PRIOR APPLICATION NUMBER: US 60/074,788

PRIOR FILING DATE: 1998-07-27

NUMBER OF SEQ ID NOS: 33142

SEQ ID NO 18164

LENGTH: 192

TYPE: PRT

ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-18164

Query Match

Best Local Similarity 100.0%; Pred. No. 2e+02;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 19 GERGQT 24  
|||||

Db 105 GERGQT 110

RESULT 44

US-09-107-532A-5858  
; Sequence 5858, Application US/09107532A  
; Patent No. 6583275

GENERAL INFORMATION:

APPLICANT: Lynn A Doucette-Stamm and David Bush

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO  
; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS

NUMBER OF SEQUENCES: 7310

ADDRESS: GENOME THERAPEUTICS CORPORATION

STREET: 100 Beaver Street

CITY: Waltham

STATE: Massachusetts

COUNTRY: USA

ZIP: 02354

COMPUTER READABLE FORM:

MEDIUM TYPE: CD-ROM ISO9660

OPERATING SYSTEM: <Unknown>

SOFTWARE: ASCII

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/107,532A

FILING DATE: 30-Jun-1998

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 60/085,598

FILING DATE: 14 May 1998

APPLICATION NUMBER: 60/051571

FILING DATE: July 2, 1997

ATTORNEY/AGENT INFORMATION:

NAME: Ariniello, Pamela Denek

REGISTRATION NUMBER: 40,489

REFERENCE/DOCKET NUMBER: GTC-012

TELECOMMUNICATION INFORMATION:

TELEPHONE: (781) 893-5007

TELEFAX: (781) 893-8277

INFORMATION FOR SEQ ID NO: 5858:

SEQUENCE CHARACTERISTICS:

LENGTH: 197 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

HYPOTHETICAL: YES

ORIGINAL SOURCE:

ORGANISM: Enterococcus faecium

FEATURE:

NAME/KEY: misc feature  
LOCATION: (B) LOCATION 1...197  
SEQUENCE DESCRIPTION: SEQ ID NO: 5858:  
US-09-107-532A-5858

Query Match 4.6%; Score 6; DB 4; Length 197;  
Best Local Similarity 100.0%; Pred. No. 2e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 86 NGQLVF 91  
Db 76 NGQLVF 81

RESULT 45  
US-09-252-991A-27431  
Sequence 27431, Application US/09252991A  
Patent No. 6551795  
GENERAL INFORMATION:  
APPLICANT: Marc J. Rubenfield et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
FILE REFERENCE: 107196.136  
CURRENT APPLICATION NUMBER: US/09/252,991A  
PRIOR FILING DATE: 1999-02-18  
PRIOR APPLICATION NUMBER: US 60/074,788  
PRIOR FILING DATE: 1998-02-18  
PRIOR APPLICATION NUMBER: US 60/094,190  
PRIOR FILING DATE: 1998-07-27  
NUMBER OF SEQ ID NOS: 33142  
SEQ ID NO 27431  
LENGTH: 199  
TYPE: PRT  
ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-27431

Query Match 4.6%; Score 6; DB 4; Length 199;  
Best Local Similarity 100.0%; Pred. No. 2e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 25 SVAPP 30  
Db 141 SVAPP 146

RESULT 46  
US-09-252-991A-26201  
Sequence 26201, Application US/09252991A  
Patent No. 6551795  
GENERAL INFORMATION:  
APPLICANT: Marc J. Rubenfield et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
FILE REFERENCE: 107196.136  
CURRENT APPLICATION NUMBER: US/09/252,991A  
PRIOR FILING DATE: 1999-02-18  
PRIOR APPLICATION NUMBER: US 60/074,788  
PRIOR FILING DATE: 1998-02-18  
PRIOR APPLICATION NUMBER: US 60/094,190  
PRIOR FILING DATE: 1998-07-27  
NUMBER OF SEQ ID NOS: 33142  
SEQ ID NO 26201  
LENGTH: 209  
TYPE: PRT  
ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-26201

Query Match 4.6%; Score 6; DB 4; Length 209;  
Best Local Similarity 100.0%; Pred. No. 2.1e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 110 RRASNG 115  
Db 110 RRASNG 115

Db 159 RRASNG 164

RESULT 47  
US-09-107-532A-4087  
Sequence 4087, Application US/09107532A  
Patent No. 6583275  
GENERAL INFORMATION:  
APPLICANT: Lynn A Doucette-Stamm and David Bush  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO  
ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS  
NUMBER OF SEQUENCES: 7310  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: GENOME THERAPEUTICS CORPORATION  
STREET: 100 Beaver Street  
CITY: Waltham  
STATE: Massachusetts  
COUNTRY: USA  
ZIP: 02354  
COMPUTER READABLE FORM:  
MEDIUM TYPE: CD-ROM ISO9660  
COMPUTER: PC  
OPERATING SYSTEM: <Unknown>  
SOFTWARE: ASCII  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/107,532A  
FILING DATE: 30-Jun-1998  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 60/085,598  
FILING DATE: 14 May 1998  
APPLICATION NUMBER: 60/051571  
FILING DATE: July 2, 1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Ariniello, Pamela Deneke  
REGISTRATION NUMBER: 40,489  
REFERENCE/DOCKET NUMBER: GTC-012  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (781)893-5007  
TELEFAX: (781)893-8277  
INFORMATION FOR SEQ ID NO: 4087:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 210 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: YES  
ORIGINAL SOURCE:  
ORGANISM: Enterococcus faecium  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (B) LOCATION 1...210  
SEQUENCE DESCRIPTION: SEQ ID NO: 4087:  
US-09-107-532A-4087

Query Match 4.6%; Score 6; DB 4; Length 210;  
Best Local Similarity 100.0%; Pred. No. 2.2e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 68 GIEIES 73  
Db 132 GIEIES 137

RESULT 48  
US-09-107-532A-5566  
Sequence 5566, Application US/09107532A  
Patent No. 6583275  
GENERAL INFORMATION:  
APPLICANT: Lynn A Doucette-Stamm and David Bush  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO  
ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS  
NUMBER OF SEQUENCES: 7310  
CORRESPONDENCE ADDRESS:  
US-09-107-532A-5566

ADDRESSEE: GENOME THERAPEUTICS CORPORATION  
STREET: 100 Beaver Street  
CITY: Waltham  
STATE: Massachusetts  
COUNTRY: USA  
ZIP: 02354

COMPUTER READABLE FORM:  
MEDIUM TYPE: CD-ROM ISO9660  
COMPUTER: PC

OPERATING SYSTEM: <Unknown>  
SOFTWARE: ASCII

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/107, 532A  
FILING DATE: 30-Jun-1998

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 60/085, 598  
FILING DATE: 14 May 1998

APPLICATION NUMBER: 60/051571  
FILING DATE: July 2, 1997

ATTORNEY/AGENT INFORMATION:  
NAME: Arinello, Pamela Deneke  
REGISTRATION NUMBER: 40,489

REFERENCE/DOCKET NUMBER: GTC-012  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (781)893-5007

TELEFAX: (781)893-8277  
INFORMATION FOR SEQ ID NO: 5566:  
SEQUENCE CHARACTERISTICS:

LENGTH: 213 amino acids  
TYPE: amino acid  
TOPOLOGY: linear

MOLECULE TYPE: protein  
HYPOTHEICAL: YES  
ORIGINAL SOURCE:

ORGANISM: Enterococcus faecium  
FEATURE:  
NAME/KEY: misc feature

LOCATION: (B) LOCATION 1...213  
SEQUENCE DESCRIPTION: SEQ ID NO: 5566:  
US-09-107-532A-5566

Query Match 4.6%; Score 6; DB 4; Length 213;  
Best Local Similarity 100.0%; Pred. No. 2.2e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 116 ETLKXI 121  
|||||  
DB 55 ETLKXI 60

RESULT 49  
US-09-252-991A-24101  
Sequence 24101, Application US/09252991A

PATENT No. 6551795  
GENERAL INFORMATION:  
APPLICANT: Marc J. Rubenfield et al.

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
FILE REFERENCE: 107196.136  
CURRENT APPLICATION NUMBER: US/09/252, 991A

PRIOR FILING DATE: 1999-02-18  
PRIOR APPLICATION NUMBER: US 60/074,788  
PRIOR FILING DATE: 1998-02-18

PRIOR APPLICATION NUMBER: US 60/094,190  
PRIOR FILING DATE: 1998-07-27  
NUMBER OF SEQ ID NOS: 33142

SEQ ID NO 24101  
LENGTH: 230

TYPE: PRT  
ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-24101

Query Match 4.6%; Score 6; DB 4; Length 230;

Best Local Similarity 100.0%; Pred. No. 2.4e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 57 ELASAV 62  
|||||  
DB 140 ELASAV 145

RESULT 50  
US-08-744-026-1

Sequence 1, Application US/08744026  
PATENT No. 5786148  
GENERAL INFORMATION:

APPLICANT: Bandman, Olga  
TITLE OF INVENTION: A NOVEL PROSTATE-SPECIFIC  
NUMBER OF SEQUENCES: 5

TITLE OF INVENTION: KALIKREIN  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: INCYTE PHARMACEUTICALS, INC.

STREET: 3174 Porter Drive  
CITY: Palo Alto  
STATE: CA

COUNTRY: US  
ZIP: 94304

COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM compatible

OPERATING SYSTEM: DOS  
SOFTWARE: FASTSEQ Version 2.0  
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/744,026  
FILING DATE: Herewith  
CLASSIFICATION: 514

PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:

ATTORNEY/AGENT INFORMATION:  
NAME: Billings, Lucy J  
REGISTRATION NUMBER: 36,749

REFERENCE/DOCKET NUMBER: PF-0154 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-855-0555

TELEFAX: 415-845-4166  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:

LENGTH: 248 amino acids  
TYPE: amino acid  
STRANDEDNESS: single

TOPOLOGY: linear  
IMMEDIATE SOURCE:  
LIBRARY:

CLONE: Consensus  
US-08-744-026-1

Query Match 4.6%; Score 6; DB 1; Length 248;  
Best Local Similarity 100.0%; Pred. No. 2.5e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 75 LGGTGA 80  
|||||  
DB 12 LGGTGA 17

Search completed: December 15, 2003, 17:07:54  
Job time : 23 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 15, 2003, 18:41:42 ; Search time 6318 seconds

(without alignments)  
10800.452 Million cell updates/sec

Title: US-09-925-301-208

Perfect score: 1668  
Sequence: 1 cacactgcctcgcgcgata.....tttgaagcttgcaggagaa 1658

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2888711 seqs, 2045481386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

GenBank1:\*  
1: gb\_ba:\*  
2: gb\_hg:\*  
3: gb\_in:\*  
4: gb\_om:\*  
5: gb\_ov:\*  
6: gb\_pat:\*  
7: gb\_ph:\*  
8: gb\_pl:\*  
9: gb\_pr:\*  
10: gb\_ro:\*  
11: gb\_seg:\*  
12: gb\_sy:\*  
13: gb\_un:\*  
14: gb\_vl:\*  
15: em\_ba:\*  
16: em\_fun:\*  
17: em\_hum:\*  
18: em\_in:\*  
19: em\_mu:\*  
20: em\_om:\*  
21: em\_or:\*  
22: em\_ov:\*  
23: em\_pat:\*  
24: em\_ph:\*  
25: em\_pl:\*  
26: em\_ro:\*  
27: em\_seg:\*  
28: em\_un:\*  
29: em\_vl:\*  
30: em\_hg\_hum:\*  
31: em\_hg\_inv:\*  
32: em\_hg\_other:\*  
33: em\_hg\_mus:\*  
34: em\_hg\_pln:\*  
35: em\_hg\_fod:\*  
36: em\_hg\_mam:\*  
37: em\_hg\_vrt:\*  
38: em\_sy:\*  
39: em\_hggo\_hum:\*  
40: em\_hggo\_mus:\*  
41: em\_hggo\_other:\*

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1549.4	92.9	1835	6	AX098215 Sequence
2	1537.4	92.2	1834	6	AX335777 Sequence
3	1537.4	92.2	1834	6	AX336076 Sequence
4	1537.4	92.2	1834	6	AX336340 Sequence
5	1537.4	92.2	1834	6	AX409567 Sequence
6	1537.4	92.2	1834	6	AX474697 Sequence
7	1537.4	92.2	1834	9	HUMPROS
8	1516.4	90.9	1809	9	BC001462
9	1475.8	88.5	1796	6	AX098193
10	1305	78.2	1726	6	AX675579
11	880.2	52.8	127769	9	AC009088
12	879.2	52.7	7008	9	HSU33446
13	877.6	52.6	195476	2	AC135044
14	857.4	51.4	882	6	AX675583
15	778.2	46.7	1142	6	AR219285
16	778.2	46.7	1169	6	AR219284
17	749	44.9	1743	10	BC003851
18	749	44.9	1856	10	AB038244
19	744.2	44.6	1797	10	AF188613
20	741.8	44.5	1161	6	AX675581
21	730.4	43.8	1753	10	AF378085
22	720	43.2	2208	10	AB017638
23	717	43.0	1864	10	AF202076
24	594.2	35.6	596	6	AX193364
25	293	17.6	185788	2	AC124461
26	293	17.6	188559	2	AC124602
27	293	17.6	219763	10	AC093175
28	291.2	17.5	5374	10	AF378086
29	287.6	17.2	232119	2	AC106629
30	287.6	17.2	278372	2	AC118182
31	223.6	13.4	1339	5	AF029404
32	220.6	13.2	1102	6	AX512287
33	220.6	13.2	1102	6	AX512289
34	207.2	12.4	944	6	AX375744
35	207.2	12.4	1613	6	AR256990
36	207.2	12.4	1613	6	AF536382
37	207.2	12.4	1755	9	BC036846
38	205.8	12.3	1130	6	AR256996
39	205.6	12.3	1606	6	AX676264
40	204.6	12.3	71177	2	AC101465
41	204.4	12.3	987	10	AF542056
42	195.8	11.7	1130	6	AR234337
43	192.8	11.6	980	6	AR253973
44	192.8	11.6	1013	9	HS4306593
45	192.8	11.6	1077	9	AY030095

#### ALIGNMENTS

RESULT 1  
AX098215 LOCUS 1835 bp DNA linear PAT 30-MAR-2001  
DEFINITION Sequence 127 from Patent WO0118542.  
ACCESSION AX098215  
VERSION AX098215.1 GI:13515326  
KEYWORDS  
SOURCE  
ORGANISM Homo sapiens (human)  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
REFERENCE 1 Lee, J., Thompson, P. and Lillie, J.  
AUTHORS Identification, assessment, prevention, and therapy of ovarian  
TITLE cancer

Pred. No. is the number of results predicted by chance to have a

**JOURNAL** Patent: WO 0118542-A 127 15-MAR-2001

Millennium Predictive Medicine, Inc. (US)

### Location/Qualifiers

source	1. .1835
--------	----------

```
/organism="Homo sapiens"
```

```
/mol_type="genomic DNA"
```

```

300 /db_xref="taxon:9606"
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040
1041
1042
1043
1044
1045
1046
1047
1048
1049
1050
1051
1052
1053
1054
1055
1056
1057
1058
1059
1060
1061
1062
1063
1064
1065
1066
1067
1068
1069
1070
1071
1072
1073
1074
1075
1076
1077
1078
1079
1080
1081
1082
1083
1084
1085
1086
1087
1088
1089
1090
1091
1092
1093
1094
1095
1096
1097
1098
1099
1100
1101
1102
1103
1104
1105
1106
1107
1108
1109
1110
1111
1112
1113
1114
1115
1116
1117
1118
1119
1120
1121
1122
1123
1124
1125
1126
1127
1128
1129
1130
1131
1132
1133
1134
1135
1136
1137
1138
1139
1140
1141
1142
1143
1144
1145
1146
1147
1148
1149
1150
1151
1152
1153
1154
1155
1156
1157
1158
1159
1160
1161
1162
1163
1164
1165
1166
1167
1168
1169
1170
1171
1172
1173
1174
1175
1176
1177
1178
1179
1180
1181
1182
1183
1184
1185
1186
1187
1188
1189
1190
1191
1192
1193
1194
1195
1196
1197
1198
1199
1200
1201
1202
1203
1204
1205
1206
1207
1208
1209
1210
1211
1212
1213
1214
1215
1216
1217
1218
1219
1220
1221
1222
1223
1224
1225
1226
1227
1228
1229
1230
1231
1232
1233
1234
1235
1236
1237
1238
1239
1240
1241
1242
1243
1244
1245
1246
1247
1248
1249
1250
1251
1252
1253
1254
1255

```

# RIGIN

Query Match 92.9%; Score 1549.4; DB 6; Length 1835;

Best Local Similarity 97.8%; Pred. No. 2e-301;

Matches 1626; Conservative 5; Mismatches 24; Indels 7; Gaps 6;

1	CAACACTGCTGCTTGGATACATC	CAGAGCGCTCCGCTTGGAGCGCGCTCCCTGCTTAAAG	60
22	CACACTGCTGCTTGGATACATC	CAAGAGCGCTCCGCTTGGAGCGCGCTCCCTGCTTAAAG	81
61	GCCAGCCTTGGACAATCTTGCTG	CCCCCTTTCAGAGCCCGGAAATTCCTTGAGATTCCTTCCCTTGAG	120
82	GCCAGCCTTGGACAATCTTGCTG	CCCCCTTTCAGAGCCCGGAAATTCCTTGAGATTCCTTCCCTTGAG	141
121	CCAAACAATCTGGGCTCGGCTT	GGACAACAACCCCAAGGCTTCTCACTTGCGGTGCTGAG	180
142	CCAAACAATCTGGGCTCGGCTT	GGACAACAACCCCAAGGCTTCTCACTTGCGGTGCTGAG	201
181	TCGCCCCAGAGGCCCCCTTGT	CTTGGGCGCATGCGCMAAGAGGGGCTCTTGGGCTTGGGCA	240
202	TCGCCCCAGAGGCCCCCTTGT	CTTGGGCGCATGCGCMAAGAGGGGCTCTTGGGCTTGGGCA	261
241	GCTGGGGGGCTGTGGGCAATTC	GTGCTCATCTTGGATTAATCTCCGGTGGGGGACAAGACGGA	300
262	GCTGGGGGGCTGTGGGCAATTC	GTGCTCATCTTGGATTAATCTCCGGTGGGGGACAAGACGGA	321
301	AGGGGCGAAGACTGCTCGGTG	GGCCCCCAAGACAAGCATCACAGTGGACGACAGTGC	360
322	AGGGGCGAAGACTGCTCGGTG	GGCCCCCAAGACAAGCATCACAGTGGACGACAGTGC	381
361	AGTGGCCGCTCAGTGGCCCTT	GGCAGGTCACAGATCACTATGAAGGGCTCATTGTGTGG	420
382	AGTGGCCGCTCAGTGGCCCTT	GGCAGGTCACAGATCACTATGAAGGGCTCATTGTGTGG	441
421	TGGCTCTCTGTGCTGAGCAG	TGGGCTGCTGAGTGTCACTGCTTCCCGACGACAGA	480
442	TGGCTCTCTGTGCTGAGCAG	TGGGCTGCTGAGTGTCACTGCTTCCCGACGACAGA	501
481	CCACAAAGAAAGCTATGAGGT	CAAGCTGGGAGGCCACACAGTGAAGCTCTTCCGAGGA	540
502	CCACAAAGAAAGCTATGAGGT	CAAGCTGGGAGGCCACACAGTGAAGCTCTTCCGAGGA	561
541	CGCCAAAGTACGACCCCTGAA	AGGACATATCCCAACCCAGGTACCTCCAGGAGGGGCTC	600
562	CGCCAAAGTACGACCCCTGAA	AGGACATATCCCAACCCAGGTACCTCCAGGAGGGGCTC	621
601	CCAGAGGCGAATTCAGTCTCC	CAACTCAGACAGACCATCACTTCCCGCTACATCCG	660
622	CCAGAGGCGAATTCAGTCTCC	CAACTCAGACAGACCATCACTTCCCGCTACATCCG	681
661	GCCCATCTGCTCCCTCGAGC	CAACCGCTCTTCCCAAGGGCTCCACATGCACCTGCAC	720
682	GCCCATCTGCTCCCTCGAGC	CAACCGCTCTTCCCAAGGGCTCCACATGCACCTGCAC	741
721	TGGCTGGGGGTATGTGGGCCC	CTCAGTGAAGGCTCTTGACGCCCAAGGCACTTGAAGCAACT	780
742	TGGCTGGGGGTATGTGGGCCC	CTCAGTGAAGGCTCTTGACGCCCAAGGCACTTGAAGCAACT	801
781	CGAGGTGCTTGTATCACTGAT	CGTGAAGCGTGTAACTGCTGTACACATCGACGCCAAGC	840
802	CGAGGTGCTTGTATCACTGAT	CGTGAAGCGTGTAACTGCTGTACACATCGACGCCAAGC	860
841	CTGAGGAGCCGCACTTGTTC	CAAGAAGACATGTGTGTGCTGTGCTATGTGAAGGGGGCA	900
861	CTGAGGAGCCGCACTTGTTC	CAAGAAGACATGTGTGTGCTGTGCTATGTGAAGGGGGCA	920

OY	901	AGGACGCTCAGCAGGGGGACTCTGAGGGGACCACTCTCTCTCCCTGATGAGGGTCTCTGT	960
Db	921	AGGACGCTCAGCAGGGTGACTCTGAGGGGCCCACTCTCTCTCTGATGAGGGTCTCTGT	980
OY	961	ACCTGACGGGCAATTGTGAGCTGGGGAGATGCTGTGGGGCCCGCAACGGCTGTGTGT	1020
Db	981	ACCTGAGGGGCAATTGTGAGCTGGGGAGATGCTGTGGGGCCCGCAACGGCTGTGTGT	1040
OY	1021	ACACTCTGGGCTCGAGTTATGCTCTCTTGATCCAAAGCAGGTGACAGAACTCCAGGCTC	1080
Db	1041	ACACTCTGGGCTCGAGTTATGCTCTCTTGATCCAAAGCAGGTGACAGAACTCCAGGCTC	1100
OY	1081	GTCGTGAGGCCCAAAACCGAGAGTCGAGGCCGACAGACAACTCTGTGAGCAGCCCTG	1140
Db	1101	GTCGTGAGGCCCAAAACCGAGAGTCGAGGCCGACAGACAACTCTGTGAGCAGCCCTG	1160
OY	1141	CCTTCAGACTCTGCCCAAGCCAGGGCTTGTGTAGAGCCCATCTTTTCTGTCTCTGGGC	1200
Db	1161	CCTTCAGACTCTGCCCAAGCCAGGGCTTGTGTAGAGCCCATCTTTTCTGTCTCTGGGC	1220
OY	1201	TGGCTCTGGGCTCTCTCTCTCCCATGCTGACGAGCACTGAGCTGGGCCCTAATTCCAGGA	1260
Db	1221	TGGCTCTGGGCTCTCTCTCTCCCATGAGGCTGACGAGCACTGAGCTGGGCCCTAATTCCAGGA	1280
OY	1261	TGATGTGATCACTCACTCAAGGACAGGAGCTGAGTCTTCCCTGATGGCTTTTGACCCAGG	1320
Db	1281	TGATGTGATCACTCACTCAAGGACAGGAGCTGAGTCTTCCCTGATGGCTTTTGACCCAGG	1340
OY	1321	GCTGACTCTTGAGCCACTCTCTCTTCAAGAACTCTGCGGAGAGCTGGGGCCCATCTTGAT	1380
Db	1341	GCTGACTCTTGAGCCACTCTCTCTTCAAGAACTCTGCGGAGAGCTGGGGCCCATCTTGAT	1400
OY	1381	CTTTGAGCCCATCTCTTCTGAGTGTCTTTTGGGACCATCACTGAGAGTACAGAGTTTAA	1440
Db	1401	CTTTGAGCCCATCTCTTCTGAGTGTCTTTTGGGACCATCACTGAGAGTACAGAGTTTAA	1460
OY	1441	CTGCTGTGAGCAATGGGCGAGGCTCTGGGCCCTTCAMCCACATGAGACAGGCCATTGGG	1500
Db	1461	CTGCTGTGAGCAATGGGCGAGGCTCTGGGCCCTTCACCCACATGAGACAGGCCATTGGC	1520
OY	1501	CGAGTCTCTGGGAG--TCTGTGGACCTTGTGTATGTAAATGAGCCCTGTGGTTCCACT	1558
Db	1521	CGAGTCTCTGGGAGGTCCTGTGGACCTTGTGGCTGTGTAAATGAGCCCTT--GGCTCCCACT	1579
OY	1559	GTTTCTGGAAGACTGTCTTCCCGGCCGCTTCCAGACTGATAGACATTTTTTTGGC	1618
Db	1580	GTTTCTGGAAGACTGTC--TCCCGGCCGCTGTGCCAGACT--GATAGACACA--TCTCTGTGC	1638
OY	1619	CNTTTCCTGTGTTTTTGGGTTGGGCAACTTTTGGAAATT	1660
Db	1637	CCTCTCCTGTGTGTCTGGGCTGGGGCCACTTTTGTGACGTT	1678

RESULT 2	AX335777	1834 bp	DNA	linear	PAT 09-JAN-2002
AX335777	LOCUS	Sequence 6286 from Patent WO0194629.			
DEFINITION	AX335777				
ACCESSION	AX335777.1	GI:18126496			
VERSION					
KEYWORDS					
SOURCE					
ORGANISM	Homo sapiens (human)				
	Homo sapiens				
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
REFERENCE					
AUTHORS	1 Young, P.E., Augustus, M., Carter, K.C., Edner, R., Endress, G.,				
	Horrigan, S., Soppet, D.R. and Weaver, Z.				
TITLE	Cancer gene determination and therapeutic screening using signature				
	gene sets				
JOURNAL	Patent: WO 0194629-A 6286 13-DEC-2001;				
	Avatlon Pharmaceuticals (US)				
FEATURES	Location/Qualifiers				

```

source      1..1834
            /organism="Homo sapiens"
            /mol_type="genomic DNA"
            /db_xref="taxon:9606"
BASE COUNT   309 a      621 c      526 g      378 t
ORIGIN
Query Match  92.2%; Score 1537.4; DB 6; Length 1834;
Best Local Similarity 97.8%; Pred. No. 5,2e-299;
Matches 1625; Conservative 5; Mismatches 24; Indels 8; Gaps 7;

QY 1 CACACGCTCCGCTTGGGATATCTCAGAGCGCTCTCCGTTGGGCGCTCCCTGCTTAGAG 60
DB |||
DB 22 CACACTGCTCGCTTGCAGATATCTCAGAGCGCTCTCCGTTGGGCGCTCCCTGCTTAGAG 81
QY 61 GCAGGCTTGGACACTTGTGCGCCCTTTCCAGCCGGGATTTCTGGGATTCCTTCCCTTAG 120
DB 82 GCACGCTTGGACACTTGTGCGCCCTTTCCAGCCGGGATTTCTGGGATTCCTTCCCTTAG 141
QY 121 CCAAGATCTGGGTCCTGCTTGCACACCAAGGCTTCTCACTTGGCTGCTGAG 180
DB 142 CCAAGATCTGGGTCCTGCTTGCACACCAAGGCTTCTCACTTGGCTGCTGAG 201
QY 181 TCTGCCCCAGGGGCGCTTGTCTTGGGCGCATGCGCMAAGAGGGGCTCTGGGCGCTGAG 240
DB 202 TCTGCCCCAGGGGCGCTTGTCTTGGGCGCATGCGCMAAGAGGGGCTCTGGGCGCTGAG 260
QY 241 GCTTGGGGGCTGTGGGCGCATCTGCTCTATTTGGAATTACTCCGGTGGGGGACAGAGGCGGA 300
DB 261 GCTTGGGGGCTGTGGGCGCATCTGCTCTATTTGGAATTACTCCGGTGGGGGACAGAGGCGGA 320
QY 301 AGGGGACAGAGGCTCTGCGGTGTGGGCGCCCAAGACCGCATCACAGTGGGACAGTGC 360
DB 321 AGGGGACAGAGGCTCTGCGGTGTGGGCGCCCAAGACCGCATCACAGTGGGACAGTGC 380
QY 361 AGTCCCGGCTCAGTGGGCGCTTGGGAGGTACAGATCACTTAAAGGCGTCATGTGTGG 420
DB 381 AGTCCCGGCTCAGTGGGCGCTTGGGAGGTACAGATCACTTAAAGGCGTCATGTGTGG 440
QY 421 TGGCTCTCTCGTGTGAGCAGTGGGTCGTGAGGTGCTCACTGTTCCCGACAGAGA 480
DB 441 TGGCTCTCTCGTGTGAGCAGTGGGTCGTGAGGTGCTCACTGTTCCCGACAGAGA 500
QY 481 CCACAGAGAGGCTTAAAGTCAAGGTGGGGGCCCAACAGTAACTCTCTACCTCCAGAGA 540
DB 501 CCACAGAGAGGCTTAAAGTCAAGGTGGGGGCCCAACAGTAACTCTCTACCTCCAGAGA 560
QY 541 GCGCAAGGTCAAGACCTTGAAGACATCATCCCCCAAGCTTACCTTCAGAGAGGCTC 600
DB 561 GCGCAAGGTCAAGACCTTGAAGACATCATCCCCCAAGCTTACCTTCAGAGAGGCTC 620
QY 601 CCAGGGCGCATTTGACCTCTCAACTCAGAGACCCCACTCTCTCCCGCTATACCTCG 660
DB 621 CCAGGGCGCATTTGACCTCTCAACTCAGAGACCCCACTCTCTCCCGCTATACCTCG 680
QY 661 GCCCATCTGCGCTCCCTGAGGACCAAGCGCTCTTCCCAAGGCGCTCACTGACGTGCAC 720
DB 681 GCCCATCTGCGCTCCCTGAGGACCAAGCGCTCTTCCCAAGGCGCTCACTGACGTGCAC 740
QY 721 TGGCTGGGGTCAATGGGCGCTCTCAGTGAAGCTCTTCAAGCCCAAGCACTGACAGCAAT 780
DB 741 TGGCTGGGGTCAATGGGCGCTCTCAGTGAAGCTCTTCAAGCCCAAGCACTGACAGCAAT 800
QY 781 CGAGGTGCTCTGATCAGTCGTGAGACGCTGAATCTGCTTGAACAATCGACGCGCAAGC 840
DB 801 CGAGGTGCTCTGATCAGTCGTGAGACGCTGAATCTGCTTGAACAATCGACGCGCAAGC 859
QY 841 CTGAAGAGCCGCACTTTGTCAAGAGAGCATGTGTGTGTCTGCTATGTGAGAGGGGCGCA 900
DB 860 CTGAAGAGCCGCACTTTGTCAAGAGAGCATGTGTGTGTCTGCTATGTGAGAGGGGCGCA 919
QY 901 AGGAGCGCTGCGAGGGGTGACTTGGGGGCGCACTTCTCTGCTTGAAGGGTCTCTGAT 960

```

```

DB 920 AGAGCGCTGCCAGGGGTGACTCTGGGGGCGCACTCTCTGCTTGTGAGAGGGTCTCTGCT 979
QY 961 ACTTGAAGGGGATTTGAGCTGGGGAGATGCTGTGGGGCGCGCAAGGCGCTGCTGCT 1020
DB |||
DB 980 ACTTGAAGGGGATTTGAGCTGGGGAGATGCTGTGGGGCGCGCAAGGCGCTGCTGCTGCT 1039
QY 1021 ACACCTTGGGCTTCAAGTATGCTCTGATCCAAAGAGGTGACAGAACTTCAAGCTC 1080
DB 1040 ACACCTTGGGCTTCAAGTATGCTCTGATCCAAAGAGGTGACAGAACTTCAAGCTC 1099
QY 1081 GTGTGTGCTCCCAAAACCAAGAGGTCCAGGCCGACAGCAACTCTGTGGACGACCTGG 1140
DB 1100 GTGTGTGCTCCCAAAACCAAGAGGTCCAGGCCGACAGCAACTCTGTGGACGACCTGG 1159
QY 1141 CCTTGAAGCTTGGCGGAGCGGCTTGTGAGGCGCACTTCTTCTGCTGCTGAGGCGC 1200
DB 1160 CCTTGAAGCTTGGCGGAGCGGCTTGTGAGGCGCACTTCTTCTGCTGCTGAGGCGC 1219
QY 1201 TGGCTCTGGGCGCTCTCTTCCCATGAGCTCAGAGCACTGAGCTGAGCCCTTCAAGGA 1260
DB 1220 TGGCTCTGGGCGCTCTCTTCCCATGAGCTCAGAGCACTGAGCTGAGCCCTTCAAGGA 1279
QY 1261 TGGATGATCACTCACTCAAGAGCAGAGCTGTGCTCTTCTGATGAGGCTTTGAGCCAGG 1320
DB 1280 TGGATGATCACTCACTCAAGAGCAGAGCTGTGCTCTTCTGATGAGGCTTTGAGCCAGG 1339
QY 1321 GCTGACTTGAAGCGACTCTCTTCTTCAAGACTCTGCGGAGGCTGGGCGCCATCTTGAT 1380
DB 1340 GCTGACTTGAAGCGACTCTCTTCTTCAAGACTCTGCGGAGGCTGGGCGCCATCTTGAT 1399
QY 1381 CTTTGAAGCCATCTTCTGAGGTGCTTTTGGAGCACTCACTGAGAGTCAAGAGTTTGA 1440
DB 1400 CTTTGAAGCCATCTTCTGAGGTGCTTTTGGAGCACTCACTGAGAGTCAAGAGTTTGA 1459
QY 1441 CTGCTGTAGCAATGAGCGAGGCTCTGCGCCCTTCACACCATGAGCCATTTGGG 1500
DB 1460 CTGCTGTAGCAATGAGCGAGGCTCTGCGCCCTTCACACCATGAGCCATTTGGG 1519
QY 1501 CGAGTCTCTGGGGAG--TCTGGGAGCCTTGTATGAAGAGCCCTGGTTCCACCT 1558
DB 1520 CGAGTCTCTGGGGAGCTCTGGGAGCCTTGTGTATGAAGAGCCCTGGTTCCACCT 1578
QY 1559 GTTCTGNAAGACTGCTTCCCGGCGGCTTCCAGACTGATGAGCACTTTTGGG 1618
DB 1579 GTTCTGNAAGACTGCTTCCCGGCGGCTTCCAGACTGATGAGCACTTCTCTGCTC 1635
QY 1619 CTTTCTCTGCTGTTTGGGTTGGGCACTTTTGGAAATT 1660
DB 1636 CTTCTCTCTGTTTCTGGGCTGGGGCGCACTTTGTGAGCTT 1677

RESULT 3
AX336076 1834 bp DNA linear PAT 09-JAN-2002
LOCUS DEFINITION Sequence 6585 from Patent WO0194629.
ACCESSION AX336076
VERSION AX336076.1 GI:18126795
KEYWORDS
SOURCE
ORGANISM Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 Young, P.E., Augustus, M., Carter, K.C., Ebner, R., Endress, G.,
Horrikan, S., Soppet, D.R. and Weaver, Z.
Cancer gene determination and therapeutic screening using signature
gene sets
Patent: WO 0194629-A 6585 13-DEC-2001;
Avalon Pharmaceuticals (US)
FEATURES
source 1..1834
/organism="Homo sapiens"
/mol_type="genomic DNA"

```

BASE COUNT 309 a 621 c 526 g 378 t  
ORIGIN

Query Match 92.2%; Score 1537.4; DB 6; Length 1834;  
Best Local Similarity 97.8%; Pred. No. 5.2e-299;  
Matches 1625; Conservative 5; Mismatches 24; Indels 8; Gaps 7;

1 CACACTCTGCTGCTGGATATCTCAGAGGCTTCCTGCGGCGCTCCCTGCTTAAAG 60  
22 CACACTGCTGCTGGATATCTCAGAGGCTTCCTGCGGCGCTCCCTGCTTAAAG 81  
61 GCGAGCTTGAACACTTGTGCGCCCTTTCAGCGCGGATTTGGATTCCTTCCCTGAG 120  
82 GCGAGCTTGAACACTTGTGCGCCCTTTCAGCGCGGATTTGGATTCCTTCCCTGAG 141  
121 CCAACATCTGGGCTCTGCTTGCACACCAACCCAGGCTTCTCACTTGCCTGAG 180  
142 CCAACATCTGGGCTCTGCTTGCACACCAACCCAGGCTTCTCACTTGCCTGAG 201  
181 TCTGCTCCAGGCGCTTGTCTGCGGCTATGCTCMAAGAGGAGTCTTGGGCTTGGCA 240  
202 TCTGCTCCAGGCGCTTGTCTGCTTCTTGGCTATGCTCMAAGAGGAGTCTTGGGCTTGGCA 260  
241 GCTGGGCGCTGAGGCTATCTGCTTATCTGATTTACTCCGATCGGGAGCAGAGAGCA 300  
261 GCTGGGCGCTGAGGCTATCTGCTTATCTGATTTACTCCGATCGGGAGCAGAGAGCA 320  
301 AGGGGCAAGACTTCTGCGGTGAGGCGCCCAAGACGATCAGAGTGGAGCAGAGTGC 360  
321 AGGGGCAAGACTTCTGCGGTGAGGCGCCCAAGACGATCAGAGTGGAGCAGAGTGC 380  
361 AGTCCCGGCTCAGTGGGCTTGGAGAGTCAAGATCACTATGAAGGGGTCCATGTTGTGG 420  
381 AGTCCCGGCTCAGTGGGCTTGGAGAGTCAAGATCACTATGAAGGGGTCCATGTTGTGG 440  
421 TGGCTCTCTGATGTCAGAGAGTGGGAGTGTGAGTGTCACTGCTTCCCGAGCAGA 480  
441 TGGCTCTCTGATGTCAGAGAGTGGGAGTGTGAGTGTCACTGCTTCCCGAGCAGA 500  
481 CCAAGAGAGGCTTATGAGGTCAAGCTGGGGGCCACACAGCTAAGTCTTACTCCAGAGA 540  
501 CCAAGAGAGGCTTATGAGGTCAAGCTGGGGGCCACACAGCTAAGTCTTACTCCAGAGA 560  
541 CCGCAAGGTCAGACCTCTGAAGACATCATCTCCCAACCCAGTACTCTCAGAGAGGCTC 600  
561 CCGCAAGGTCAGACCTCTGAAGACATCATCTCCCAACCCAGTACTCTCAGAGAGGCTC 620  
601 CCGAGGCGACATTTGACCTCTCCAACTCAGAGAGCCATCACTTCCCGCTACATCCG 660  
621 CCGAGGCGACATTTGACCTCTCCAACTCAGAGAGCCATCACTTCCCGCTACATCCG 680  
661 GCCCATCTGCTCTCCGAGCACAACGCTCTTCCCAACGGGCTCCACTGACCTGTCACT 720  
681 GCCCATCTGCTCTCCGAGCACAACGCTCTTCCCAACGGGCTCCACTGACCTGTCACT 740  
721 TGGCTGGGCTCAGTGTGGCCCCCTCAGTGAAGCTCTCTGACGCCCAAGCACTGAGCAACT 780  
741 TGGCTGGGCTCAGTGTGGCCCCCTCAGTGAAGCTCTCTGACGCCCAAGCACTGAGCAACT 800  
781 CAGAGGCTCTGATCAGTGTGAGAGAGTGTGATCTGCTGATCAACATGACGCGCAAGC 840  
801 CAGAGGCTCTGATCAGTGTGAGAGAGTGTGATCTGCTGATCAACATGACGCGCAAGC 860  
841 CTGAGAGGCGGCACTTGTGCTCAAGAGAGAGTGTGATCTGCTGATCAACATGACGCGGCA 900  
860 CTGAGAGGCGGCACTTGTGCTCAAGAGAGAGTGTGATCTGCTGATCAACATGACGCGGCA 919  
901 AGGAGCGCTTGCAGAGGAGTCTTGGGGGCCCACTTCTGCTGCTGAGAGGCTTCTGCT 960  
920 AGGAGCGCTTGCAGAGGAGTCTTGGGGGCCCACTTCTGCTGCTGAGAGGCTTCTGCT 979  
961 ACCTGACGGGCACTTGTAGGCTGGGGAGATGCTGTGGGGGCCCGCAAGGCTGTGTGT 1020

|||||  
980 ACCTGACGGGCACTGTGAGCTGGGGAGATGCTGTGGGGCCCGCAACAGGCTGTGTGT 1039  
1021 ACACCTGAGCTCCAGGCTATGCTTCTGATCCAAAGAGTGAACAATCCAGGCTC 1080  
1040 ACACCTGAGCTCCAGGCTATGCTTCTGATCCAAAGAGTGAACAATCCAGGCTC 1099  
1081 GTGTGATGCCCAACCCAGAGAGTCCAGGCCGACAGCAACTCTGTGTGAGCAGCACTGTG 1140  
1100 GTGTGATGCCCAACCCAGAGAGTCCAGGCCGACAGCAACTCTGTGTGAGCAGCACTGTG 1159  
1141 CTTGAGCTTGTGCTCCAGGCCCAAGGCTTGTGAGGCCATCTTTTCTGCTTGGGCTC 1200  
1160 CTTGAGCTTGTGCTCCAGGCCCAAGGCTTGTGAGGCCATCTTTTCTGCTTGGGCTC 1219  
1201 TGGCTTGTGGGCTCTCTTCCCATGAGTCAAGAGCACTGAGCTGAGCCCTTACCTCAAGA 1260  
1220 TGGCTTGTGGGCTCTCTTCCCATGAGTCAAGAGCACTGAGCTGAGCCCTTACCTCAAGA 1279  
1261 TGGATGATCACTCAAGAGCAGAGAGCTGCTTCCCTGATAGGCTTGTGAGCCAGG 1320  
1280 TGGATGATCACTCAAGAGCAGAGAGCTGCTTCCCTGATAGGCTTGTGAGCCAGG 1339  
1321 GCTTGAATTTAGGCACTCTTCTTCAAGACTCTGCGGAGGCTTGGGCTCCATCTTGAAT 1380  
1340 GCTTGAATTTAGGCACTCTTCTTCAAGACTCTGCGGAGGCTTGGGCTCCATCTTGAAT 1399  
1381 CTTTGAAGCCATTTCTTGGGAGTCTTTTGGGACATCACTGAGAGTCAAGAGTTTGA 1440  
1400 CTTTGAAGCCATTTCTTGGGAGTCTTTTGGGACATCACTGAGAGTCAAGAGTTTGA 1459  
1441 CTGCTGTAGCAATGAGGAGGAGGCTCTGAGCCCTTCAAGCAGCAGTGAAGGCTTGA 1500  
1460 CTGCTGTAGCAATGAGGAGGAGGCTCTGAGCCCTTCAAGCAGCAGTGAAGGCTTGA 1519  
1501 CGAGNTCTGGGAGG-TCCTGGAACCTTGTGTAATAAGAGGCTTGGGCTTCCAGCT 1558  
1520 CGAGNTCTGGGAGGCTCTGAGGAGGCTTGTGTAATAAGAGGCTTGGGCTTCCAGCT 1578  
1559 GTTTCCTGGAAGCTGCTTCCCGGCGGCTTCCAGGCTTGTGTAATAAGAGGCTTGGGCTTCCAGCT 1618  
1579 GTTTCCTGGAAGCTGCTTCCCGGCGGCTTCCAGGCTTGTGTAATAAGAGGCTTGGGCTTCCAGCT 1635  
1619 CTTTCCCTGCTGTTTGGGCTTGGGCACTTTTGGAGTTT 1660  
1636 CTTTCCCTGCTGTTTGGGCTTGGGCACTTTTGGAGCTT 1677

RESULT 4  
AX336340 1834 bp DNA linear PAT 09-JAN-2002  
LOCUS  
DEFINITION Sequence 6849 from Patent WO0194629.  
ACCESSION AX336340  
VERSION AX336340.1 GI:18127059  
KEYWORDS  
SOURCE  
ORGANISM Homo sapiens (human)  
REFERENCE  
AUTHORS Young, P.E., Augustus, M., Carter, K.C., Ebner, R., Endress, G.,  
Horrigan, S., Soppet, D.R. and Weaver, Z.  
TITLE Cancer gene determination and therapeutic screening using signature  
gene sets  
JOURNAL Patent: WO 0194629-A 6849 13-DEC-2001;  
FEATURES  
source  
1. .1834  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
BASE COUNT 309 a 621 c 526 g 378 t  
ORIGIN

Query Match 92.2%; Score 1537.4; DB 6; Length 1834;  
 Best Local Similarity 97.8%; Pred. No. 5.2e-299;  
 Matches 1625; Conservative 5; Mismatches 24; Indels 8; Gaps 7;

1 CACACGCTGCTGCTGGATGATCTCAGAGGCTCTCCGTTGGGCGCTCCCTGCTTAGAG 60  
 22 CACACGCTGCTGCTGGATGATCTCAGAGGCTCTCCGTTGGGCGCTCCCTGCTTAGAG 81  
 61 GCCAGCCTTGAGACATCTGCGCCCTTTCAGCCGGATTCGGATTCCTCCCTGAG 120  
 82 GCCAGCCTTGAGACATCTGCGCCCTTTCAGCCGGATTCGGATTCCTCCCTGAG 141  
 121 CCAACATCTGGGCTCTGCTGCAACAACCCCAAGGCTTCTTACCTTGCGCTGAG 180  
 142 CCAACATCTGGGCTCTGCTGCAACAACCCCAAGGCTTCTTACCTTGCGCTGAG 201  
 181 TCTGCCCCGAGGGGCTTGTCTGCGGCAATGCGCMAAGAGGGGCTCTGAG 240  
 202 TCTGCCCCGAGGGGCTTGTCTGCGGCAATGCGCMAAGAGGGGCTCTGAG 260  
 241 GCTGGGGGCTGAGGCTTCTGCTGATCTGATGATTCCTCCGTTGGGCGCTGAG 300  
 261 GCTGGGGGCTGAGGCTTCTGCTGATCTGATGATTCCTCCGTTGGGCGCTGAG 320  
 301 AGGGGAGAGAGCTCTGCGGCTGCGCCCAAGACGCACTCAGAGTGGAGAGTGC 360  
 321 AGGGGAGAGAGCTCTGCGGCTGCGCCCAAGACGCACTCAGAGTGGAGAGTGC 380  
 361 AGTCCCGGCTCAGTGGGCTGCGGAGTCAAGCATCACTTAAGAGGCTCATGTGTG 420  
 381 AGTCCCGGCTCAGTGGGCTGCGGAGTCAAGCATCACTTAAGAGGCTCATGTGTG 440  
 421 TGGCTCTCTGATGCTGAGAGAGTGGTGTGCTGCTGCTGCTGCTGCTGCTGCTG 480  
 441 TGGCTCTCTGATGCTGAGAGAGTGGTGTGCTGCTGCTGCTGCTGCTGCTGCTG 500  
 481 CCACAAGAGAGCTTGAAGTCAAGTGGGGGCCCAAGCACTGACTCCAGAG 540  
 501 CCACAAGAGAGCTTGAAGTCAAGTGGGGGCCCAAGCACTGACTCCAGAG 560  
 541 CGCCAAAGTCAAGACCTTGAAGATCATTCCTCCCAAGCTTCTCAGAGAGGCTC 600  
 561 CGCCAAAGTCAAGACCTTGAAGATCATTCCTCCCAAGCTTCTCAGAGAGGCTC 620  
 601 CGAGGGGCAATTCGACTCTCCATCTCAGAGACCCATCACTTCTCCGCTTACATTCG 660  
 621 CGAGGGGCAATTCGACTCTCCATCTCAGAGACCCATCACTTCTCCGCTTACATTCG 680  
 661 GCCCATCTGCTCTCCCTGAGCAAGCGCTCTTCCCAAGGCTTCACTGCACTGTAC 720  
 681 GCCCATCTGCTCTCCCTGAGCAAGCGCTCTTCCCAAGGCTTCACTGCACTGTAC 740  
 721 TGGCTGGGGTCAATGTGGGCTCTCAGTGAAGCTTCTGAGCCCAAGCACTGAGCACT 780  
 741 TGGCTGGGGTCAATGTGGGCTCTCAGTGAAGCTTCTGAGCCCAAGCACTGAGCACT 800  
 781 CGAGGGCTCTGATCAGTGTGAGAGCTGATGCTGCTGATCAACATCGACGCAAGC 840  
 801 CGAGGGCTCTGATCAGTGTGAGAGCTGATGCTGCTGATCAACATCGACGCAAGC 859  
 841 CTGAGAGAGCGCACTTCTGCAAGAGACATGATGCTGCTGATGAGAGGAGCA 900  
 860 CTGAGAGAGCGCACTTCTGCAAGAGACATGATGCTGCTGATGAGAGGAGCA 919  
 901 AGGAGCGCTGCTGCAAGGATGATCTTGGGGGCCCACTCTCTGCTGAGAGGATCTCTG 960  
 920 AGGAGCGCTGCTGCAAGGATGATCTTGGGGGCCCACTCTCTGCTGAGAGGATCTCTG 979  
 961 ACCTGAGCGGCAATGATGAGTGGGAGATGCTTGGGGGCCCAAGAGGCTGATGAT 1020  
 980 ACCTGAGCGGCAATGATGAGTGGGAGATGCTTGGGGGCCCAAGAGGCTGATGAT 1039

1021 ACACCTGCGCTCCAGCTATGCTCTCTGATGATCAAGAGGATGACAGAACTCCAGCCTC 1080  
 1040 ACACCTGCGCTCCAGCTATGCTCTCTGATGATCAAGAGGATGACAGAACTCCAGCCTC 1099  
 1081 GTGTGATGCGCCCAAAACCCAGAGAGTCCAGAGCTCCAGAGCACTCTGTGAGAGCTG 1140  
 1100 GTGTGATGCGCCCAAAACCCAGAGAGTCCAGAGCTCCAGAGCACTCTGTGAGAGCTG 1159  
 1141 CCTGAGCTGCTCCAGAGAGGCTGCTGAGAGGCTCCAGAGCACTCTGTGAGAGCTG 1200  
 1160 CCTGAGCTGCTCCAGAGAGGCTGCTGAGAGGCTCCAGAGCACTCTGTGAGAGCTG 1219  
 1201 TGGCTGAGGCTCTCTCTCCAGAGGCTGCTGAGAGGCTCCAGAGCACTCTGTGAGAG 1260  
 1220 TGGCTGAGGCTCTCTCTCCAGAGGCTGCTGAGAGGCTCCAGAGCACTCTGTGAGAG 1279  
 1261 TGGATGATCAGATCTGAGAGAGAGAGGCTGCTCTCTGATGAGGCTTGTGAGAGCCAG 1320  
 1280 TGGATGATCAGATCTGAGAGAGAGAGGCTGCTCTCTGATGAGGCTTGTGAGAGCCAG 1339  
 1321 GCTGACTTGGAGGCACTGCTCTCTCAGAGACTGCGGAGGCTGAGGCTCCATCTTAT 1380  
 1340 GCTGACTTGGAGGCACTGCTCTCTCAGAGACTGCGGAGGCTGAGGCTCCATCTTAT 1399  
 1381 CTTTGAAGCCATCTCTGAGGCTGCTCTTGTGAGAGCACTGAGAGTCAAGAGTTTAT 1440  
 1400 CTTTGAAGCCATCTCTGAGGCTGCTCTTGTGAGAGCACTGAGAGTCAAGAGTTTAT 1459  
 1441 CTGCTGTGAGCAATGAGAGAGGCTCTGAGGCTTCAAGCACTGAGAGTCAAGAGTTTAT 1500  
 1460 CTGCTGTGAGCAATGAGAGAGGCTCTGAGGCTTCAAGCACTGAGAGTCAAGAGTTTAT 1519  
 1501 CGAGTCTCGGGGAG--TCTGAGGAGCTTGTGATGAAGAGGCTGAGGCTTCCAGCT 1558  
 1520 CGAGTCTCGGGGAGCTCTGAGGAGCTTGTGATGAAGAGGCTTCCAGCT 1578  
 1559 GTTCTGAGAGAGCTGCTTCCAGGCTTGTGAGGCTTCCAGAGTGAAGAGCACTTTTGTGC 1618  
 1579 GTTCTGAGAGAGCTGCTTCCAGGCTTGTGAGGCTTCCAGAGTGAAGAGCACTTTTGTGC 1635  
 1619 CTTTCCCTGCTGCTTGTGAGGCTTGTGAGGCTTGTGAGGCTTGTGAGGCTTGTGAG 1660  
 1636 CTTTCCCTGCTGCTTGTGAGGCTTGTGAGGCTTGTGAGGCTTGTGAGGCTTGTGAG 1677

RESULT 5  
 AX409567 1834 bp DNA linear PAT 14-JUN-2002  
 LOCUS AX409567  
 DEFINITION Sequence 2214 from Patent WO0229103.  
 ACCESSION AX409567  
 VERSION AX409567.1 GI:21442272  
 KEYWORDS  
 SOURCE  
 ORGANISM Homo sapiens (human)  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE  
 1 Alvarres, C., Horne, D., Peres-da-Silva, S. and Vockley, J. G.  
 Gene expression profiles in liver cancer  
 Patent: WO 0229103-A 2214 11-APR-2002;  
 JOURNAL GENE LOGIC INC (US)  
 FEATURES  
 source 1..1834  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"  
 /note="EMBL/GenBank Accession No. I41351"  
 BASE COUNT 309 a 621 c 526 g 378 t  
 ORIGIN

Query Match 92.2%; Score 1537.4; DB 6; Length 1834;  
 Best Local Similarity 97.8%; Pred. No. 5.2e-299;  
 Matches 1625; Conservative 5; Mismatches 24; Indels 8; Gaps 7;

QY 1 CACACTGCTCGCTTGGGATTAATCAGAGCCTCCGCTTGCGGCTCCCTGCTTAAG 60  
 Db 22 CACACTGCTCGCTTGGGATTAATCAGAGCCTCCGCTTGCGGCTCCCTGCTTAAG 81  
 QY 61 GCCAGCCTTGGACACTTGTGCTGCTTTCAGGCCGGAATCTTGAGTCTTCTCTGAG 120  
 Db 82 GCCAGCCTTGGACACTTGTGCTGCTTTCAGGCCGGAATCTTGAGTCTTCTCTGAG 141  
 QY 121 CCACATCTGGGCTCTGCTTGCACACACCCAGAGCTTCTTACCTTGCTGCTGAG 180  
 Db 142 CCACATCTGGGCTCTGCTTGCACACACCCAGAGCTTCTTACCTTGCTGCTGAG 201  
 QY 181 TCGGCCCGAGGGGCTTGTCTGGGCGATGGCGMAAGAGGGGCTCTGGGGCTGGGCA 240  
 Db 202 TCGGCCCGAGGGGCTTGTCTGGGCGATGGCGMAAGAGGGGCTCTGGGGCTGGGCA 260  
 QY 241 GCTGGGGGCTGTGGGCAATCTGCTTATCTTGATTAATCTCGGTCGGGGACAGAGCGGA 300  
 Db 261 GCTGGGGGCTGTGGGCAATCTGCTTATCTTGATTAATCTCGGTCGGGGACAGAGCGGA 320  
 QY 301 AGGGGCAAGAGCTTCTGCGGTGTGCTCCCAAGCAGCATCAAGGTGGACAGCTGC 360  
 Db 321 AGGGGCAAGAGCTTCTGCGGTGTGCTCCCAAGCAGCATCAAGGTGGACAGCTGC 380  
 QY 361 AGTGGCGGCTGAGTGGCGCTGGGAGGTCAGATCACTATGAAGGGGTCATGTGTGG 420  
 Db 381 AGTGGCGGCTGAGTGGCGCTGGGAGGTCAGATCACTATGAAGGGGTCATGTGTGG 440  
 QY 421 TGGCTCTCTGTCTGAGAGAGTGGGTGTGTGAGCTGCTCACTGCTTCCCAAGCAGCA 480  
 Db 441 TGGCTCTCTGTCTGAGAGAGTGGGTGTGTGAGCTGCTCACTGCTTCCCAAGCAGCA 500  
 QY 481 CCAAGAGAGAGCTTATGAGGTCAAGCTGGGGGCTCCACAGCTGATCTTCTTCCAGGA 540  
 Db 501 CCAAGAGAGAGCTTATGAGGTCAAGCTGGGGGCTCCACAGCTGATCTTCTTCCAGGA 560  
 QY 541 CGCAAGGTGAGGACCTGAAAGCATCATCCCAAGCTCACTCCAGAGAGGCTC 600  
 Db 561 CGCAAGGTGAGGACCTGAAAGCATCATCCCAAGCTCACTCCAGAGAGGCTC 620  
 QY 601 CCAAGGCGCATTTGACTCTCTCAACTCAGACAGCCATCACTTCCCGCTACATCCG 660  
 Db 621 CCAAGGCGCATTTGACTCTCTCAACTCAGACAGCCATCACTTCCCGCTACATCCG 680  
 QY 661 GCCCATCTGCTCTCCCTGAGCAGCAAGCTTCTTCCCAAGGCTCCACTGCTGCTCAC 720  
 Db 681 GCCCATCTGCTCTCCCTGAGCAGCAAGCTTCTTCCCAAGGCTCCACTGCTGCTCAC 740  
 QY 721 TGGCTGGGGTCAATGTGGGCTCCCTCAGAGAGCTCTGAGCCCAAGCAGCATGCAAGT 780  
 Db 741 TGGCTGGGGTCAATGTGGGCTCCCTCAGAGAGCTCTGAGCCCAAGCAGCATGCAAGT 800  
 QY 781 CGAGGCTCTCTGATCAGTGTGAGAGCTGTGTAATCTCTGTAACATGACGCGCAAGC 840  
 Db 801 CGAGGCTCTCTGATCAGTGTGAGAGCTGTGTAATCTCTGTAACATGACGCGCAAGC 859  
 QY 841 CTGAGAGGCGGCACTTTGTCCAAGAGCATGTGTGTCTGCTTATGTGAGGGGGCA 900  
 Db 860 CTGAGAGGCGGCACTTTGTCCAAGAGCATGTGTGTCTGCTTATGTGAGGGGGCA 919  
 QY 901 AGGAGCGCTGCAAGGTGAGCTCTGGGGGCCCACTCTGCGCTGTGAGAGGGTCTGGT 960  
 Db 920 AGGAGCGCTGCAAGGTGAGCTCTGGGGGCCCACTCTGCGCTGTGAGAGGGTCTGGT 979  
 QY 961 ACCTGACGGGCAATGTAGCTGGGAGATGCTGTGGGGCCCGCAAGAGGCTGTGTGT 1020  
 Db 980 ACCTGACGGGCAATGTAGCTGGGAGATGCTGTGGGGCCCGCAAGAGGCTGTGTGT 1039  
 QY 1021 ACACCTGGGCTCAAGCTATGCTCTCTGATTCAAAGCAGGTACAGAACTCCAGCTTC 1080  
 Db 1040 ACACCTGGGCTCAAGCTATGCTCTCTGATTCAAAGCAGGTACAGAACTCCAGCTTC 1099

QY 1081 GTGTGATGCCCCMAACCAAGAGTCCAGCCCGACAGCAACTGTGTGGACCACTTGG 1140  
 Db 1100 GTGTGATGCCCCMAACCAAGAGTCCAGCCCGACAGCAACTGTGTGGACCACTTGG 1159  
 QY 1141 CCTTGAAGCTTGGCCCCAGGCGCTGTGAGGCTCATCTTTTCTGCTCTGGGCT 1200  
 Db 1160 CCTTGAAGCTTGGCCCCAGGCGCTGTGAGGCTCATCTTTTCTGCTCTGGGCT 1219  
 QY 1201 TGGCTTGGGCTCCCTCTCCCATGAGTCTGAGGAGCACTGAGCTGAGCTTCCAGGA 1260  
 Db 1220 TGGCTTGGGCTCCCTCTCCCATGAGTCTGAGGAGCACTGAGCTGAGCTTCCAGGA 1279  
 QY 1261 TGGATGATCACAATCAAGAGCAGAGCTGTGCTTCTGATGAGCTTGGACCCAG 1320  
 Db 1280 TGGATGATCACAATCAAGAGCAGAGCTGTGCTTCTGATGAGCTTGGACCCAG 1339  
 QY 1321 GCTGATGAGCTCACTCTTCTTCAAGACTCTGCGGAGGCTGGGGCCCATCTTGTAT 1380  
 Db 1340 GCTGATGAGCTCACTCTTCTTCAAGACTCTGCGGAGGCTGGGGCCCATCTTGTAT 1399  
 QY 1381 CTTTGAGGCCATTTCTTGGGATGTGCTTTTGGAGCATCATGAGAGTCAGAGTTTAA 1440  
 Db 1400 CTTTGAGGCCATTTCTTGGGATGTGCTTTTGGAGCATCATGAGAGTCAGAGTTTAA 1459  
 QY 1441 CTGCTGTGACATGAGCCAGAGCTTGTGCTCCCTCAGCCATGAGACCCATGGG 1500  
 Db 1460 CTGCTGTGACATGAGCCAGAGCTTGTGCTCCCTCAGCCATGAGACCCATGGG 1519  
 QY 1501 CGAGTCTCTGGGAG--TCTTGGAGCTTGTGTATGAATAAGGCTTGGTCCACT 1558  
 Db 1520 CGAGTCTCTGGGAGCTTGTGAGACCTTGTGTATGAATAAGGCTTGGTCCACT 1578  
 QY 1559 GTTCTTGAAGACTGCTTCCGCGCCGCTTCCAGACTGATGAGCACTTTTGTTC 1618  
 Db 1579 GTTCTTGAAGACTGCTTCCGCGCCGCTTCCAGACTGATGAGCACTTTTGTTC 1635  
 QY 1619 CMTTCCCTGTGTGTGTGGGTTGGGCACTTTTGGAGTTT 1660  
 Db 1636 CMTTCCCTGTGTGTGTGGGCTGGGCGCACCTTTGTGACACTT 1677

RESULT 6  
 AX474697  
 LOCUS AX474697 1834 bp DNA linear PAT 12-AUG-2002  
 DEFINITION Sequence 1 from Patent WO0221133.  
 ACCESSION AX474697  
 VERSION AX474697.1 GI:22214031  
 KEYWORDS  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1  
 Mok, S.C. and Wong, K.K.  
 Methods of detecting cancer based on prostaasin  
 Patent: WO 0221133-A 1 14-MAR-2002;  
 THE BRIGHAM AND WOMEN'S HOSPITAL, INC. (US)  
 FEATURES  
 source  
 1. .1834  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"  
 229..1260  
 /note="unnamed protein product"  
 /codon\_start=1  
 /protein\_id="CAD43672.1"  
 /db\_xref="GI:22214032"  
 /translation="MAOKGVLPGRGOLGAVAILLYLGLSLRSGAEAGAPCGVAPOL  
 HODLSISEDAKSTLMDIIPHSYLOBSQGDALLOLSRPIFSFYIRPICAPANA  
 SPENGLHCTVGTGAVPSVSLTPKPLQOLEVPLISRETCLNLYIDAKPEPHVQ  
 EDWVCAGYVEGGDADCOGDSGGPLSCPVEGLWYLTGIVMGDAGCARNPBGVYTLAS  
 YSMWIOSKVELQPRVVPOTOBSPDNLGSHLAFSSAPAOGLRPIILPLGLALG



BASE COUNT 309 a 621 c 526 g 378 t  
 ORIGIN L1SPMISEH"  
 Query Match 92.2%; Score 1537.4; DB 6; Length 1834;  
 Best Local Similarity 97.8%; Pred. No. 5.2e-299;  
 Matches 1655; Conservative 5; Mismatches 24; Indels 8; Gaps 7;

```

QY 1 CACACTGCTGCTCGGATATCTCAGGCGCTCTCCGTTGGGCGCGCTCTGCTTAGAG 60
DB 22 CACACTGCTGCTCGGATATCTCAGGCGCTCTCCGTTGGGCGCGCTCTGCTTAGAG 81
QY 61 GCCAGCTTGGACACTTCTGCTCCCTTTCCAGCCCGGATTTGGGATCTTCTCTGAG 120
DB 82 GCCAGCTTGGACACTTCTGCTCCCTTTCCAGCCCGGATTTGGGATCTTCTCTGAG 141
QY 121 CCAACATCTGGGCTGCTGCTTGGACACACCCCAAGGCTTCCATCCTTGGTGGCGAG 180
DB 142 CCAACATCTGGGCTGCTGCTTGGACACACCCCAAGGCTTCCATCCTTGGTGGCGAG 201
QY 181 TCTGCCCGAGGGGCGCTTGTCTGAGGCAATGCGCMAAGAGGGGCTCTGGGGCTGAG 240
DB 202 TCTGCCCGAGGGGCGCTTGTCTGAGGCAATGCGCMAAGAGGGGCTCTGGGGCTGAG 260
QY 241 GCTGGGGGCTGTGGGCAATCTGCTTATCTTGGATTACTCGGCTGGGGACAGAGCCAG 300
DB 261 GCTGGGGGCTGTGGGCAATCTGCTTATCTTGGATTACTCGGCTGGGGACAGAGCCAG 320
QY 301 AGGGGAGAGAGCTTCTGCGGTGTGGCCCCCGAGACAGATCAAGGTGGAGAGCTGC 360
DB 321 AGGGGAGAGAGCTTCTGCGGTGTGGCCCCCGAGACAGATCAAGGTGGAGAGCTGC 380
QY 361 AGTCCCGGCTAGTGGGCGCTTGGAGGTCAGCATCACTATGAAGAGGCTCATGTGTG 420
DB 381 AGTCCCGGCTAGTGGGCGCTTGGAGGTCAGCATCACTATGAAGAGGCTCATGTGTG 440
QY 421 TGGCTCTCTGTGTCTGAGAGAGTGGGTGTGTGAGTGTCTCACTGTTCCCGAGCA 480
DB 441 TGGCTCTCTGTGTCTGAGAGAGTGGGTGTGTGAGTGTCTCACTGTTCCCGAGCA 500
QY 481 CCAAGAGAGAGCTTATGAGGTCAAGCTGGGGGCGCCACAGCTAGATCTCTACCCAG 540
DB 501 CCAAGAGAGAGCTTATGAGGTCAAGCTGGGGGCGCCACAGCTAGATCTCTACCCAG 560
QY 541 CCGCAAGGTCAAGACCTTGAAGAGCATCACTCCCGACCCAGCTTACCTCAGAGGGCTC 600
DB 561 CCGCAAGGTCAAGACCTTGAAGAGCATCACTCCCGACCCAGCTTACCTCAGAGGGCTC 620
QY 601 CCAAGGCGCATTTGCACTCTCTCCACTCAGACAGACCCATCACTTCTCCGCTACATCCG 660
DB 621 CCAAGGCGCATTTGCACTCTCTCCACTCAGACAGACCCATCACTTCTCCGCTACATCCG 680
QY 661 GCCCATCTGCTCTCTCTGAGCGCAAGCGCTCTTCCCAAGCGCTTCACTGCACTGTAC 720
DB 681 GCCCATCTGCTCTCTCTGAGCGCAAGCGCTCTTCCCAAGCGCTTCACTGCACTGTAC 740
QY 721 TGGCTGGGGTCAATGTGGGCGCGCTCAGTGAAGCTCTGAGGCGCCAGCACTGAGAGACT 780
DB 741 TGGCTGGGGTCAATGTGGGCGCGCTCAGTGAAGCTCTGAGGCGCCAGCACTGAGAGACT 800
QY 781 CGAGGTGCTCTGATCAGTGTGAGAGAGTGTAACTGCTGTACAACTGACGCGCAAGC 840
DB 801 CGAGGTGCTCTGATCAGTGTGAGAGAGTGTAACTGCTGTACAACTGACGCGCAAGC 859
QY 841 CTGAGAGAGCGGCACTTTGTCCAAAGAGAGAGATGTGTGTGTGCTATGTGAGAGGGGCA 900
DB 860 CTGAGAGAGCGGCACTTTGTCCAAAGAGAGAGATGTGTGTGTGCTATGTGAGAGGGGCA 919
QY 901 AGGAGCGCTGCGAGGCTGCTGAGGGGCGCACTCTGCGCTGTGAGAGGGGCTCTGGT 960
DB 920 AGGAGCGCTGCGAGGCTGCTGAGGGGCGCACTCTCTGCGCTGTGAGAGGGGCTCTGGT 979
QY 961 ACCTGACGGGCAATTGAGACTGGGAGAGATGCTGTGGGGCGCGCAAGGCGCTGTGTGT 1020

```

```

DB 980 ACCTGACGGGCAATTGAGACTGGGAGAGATGCTGTGGGGCGCGCAAGGCGCTGTGTGT 1039
QY 1021 ACACCTGAGCTCCAGACTATGCTCTCTGAGATCCAAAGAGAGGACAGAACTCCAGCCTC 1080
DB 1040 ACACCTGAGCTCCAGACTATGCTCTCTGAGATCCAAAGAGAGGACAGAACTCCAGCCTC 1099
QY 1081 GTGTGGTGGCCCAAAACCCAGAGTCCAGAGCCCGACAGCAACTCTGTGGACCACTCTGG 1140
DB 1100 GTGTGGTGGCCCAAAACCCAGAGTCCAGAGCCCGACAGCAACTCTGTGGACCACTCTGG 1159
QY 1141 CCTTACGCTCTGCCCCAGCTCAGGGCTTGTCTGAGGCCATCTTTTCTGCTCTGGGCC 1200
DB 1160 CCTTACGCTCTGCCCCAGCTCAGGGCTTGTCTGAGGCCATCTTTTCTGCTCTGGGCC 1219
QY 1201 TGGCTCTGGGCTCTGCTCTCCATGGCTCAGAGAGACTGAGCTGGCCCTACTCTCAGAGA 1260
DB 1220 TGGCTCTGGGCTCTGCTCTCCATGGCTCAGAGAGACTGAGCTGGCCCTACTCTCAGAGA 1279
QY 1261 TGGATGATCAGACTCAAGAGAGAGAGCTGTGCTCTCTGATGGCTTTGAGCCAGG 1320
DB 1280 TGGATGATCAGACTCAAGAGAGAGAGCTGTGCTCTCTGATGGCTTTGAGCCAGG 1339
QY 1321 GCTTGACTTGAAGCACTCTCTCTCTTCAAGACTCTGCGGAGGCTGGGGCCCATCTGTAT 1380
DB 1340 GCTTGACTTGAAGCACTCTCTCTCTTCAAGACTCTGCGGAGGCTGGGGCCCATCTGTAT 1399
QY 1381 CTTTGAAGCCATCTTCTGAGGTGTCTTTTGGAGCACTGAGAGAGTCAAGAGTTTA 1440
DB 1400 CTTTGAAGCCATCTTCTGAGGTGTCTTTTGGAGCACTGAGAGAGTCAAGAGTTTA 1459
QY 1441 CTGCTGTGAGCAATGAGGCGAGAGCTCTGGGCGCTTCCAGCACCAGTGAAGCCCATTTG 1500
DB 1460 CTGCTGTGAGCAATGAGGCGAGAGCTCTGGGCGCTTCCAGCACCAGTGAAGCCCATTTG 1519
QY 1501 CGAGTCTCTGGGAGG-TCCTGGAGCTTTGGTATGAAGATGAGCTTGGTCCACCT 1558
DB 1520 CGAGTCTCTGGGAGGCTCTGGGAGCCTTGGCTATGAAGATGAGCCCT-GGCTCCACCT 1578
QY 1559 GTTTCCTGAGAGCTGCTCTTCCCGGCGCGCTTCCAGCACTGATGAGCACTTTTTC 1618
DB 1579 GTTTCCTGAGAGCTGCTCTTCCCGGCGCGCTTCCAGCACTGATGAGCACTTTTTC 1635
QY 1619 CTTTCCCTGCTGTTTTGGGTTGGGCACTTTTGAAGTTT 1660
DB 1636 CTTTCCCTGCTGTGTCTGGGCTGGGCGCACCTTTGTGACCTT 1677

```

RESULT 7  
 HUMPROS 1834 bp mRNA linear PRI 15-JUN-1995  
 LOCUS Homo sapiens prostatein mRNA, complete cds.  
 DEFINITION L41351  
 ACCESSION L41351.1 GI:862304  
 VERSION prostatein.  
 KEYWORDS Homo sapiens (human)  
 SOURCE  
 ORGANISM  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE  
 1 (bases 1 to 1834)  
 Yu J.X., Chao L. and Chao J.  
 Molecular cloning, tissue-specific expression, and cellular  
 localization of human prostatein mRNA (1995)  
 J. Biol. Chem. 270 (22), 13483-13489  
 JOURNAL  
 MEDLINE  
 PUBMED  
 COMMENT  
 FEATURES  
 source  
 1..1834  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /strain="Caucasian"  
 /db\_xref="taxon:9606"

```

mRNA
5'UTR
CDS
<1..>1834
<1..>228
229..1260
/codon_start=1
/product="prolactin"
/db_xref="GI:862305"
/translation="MAQKGLGPGOLGAVAILLYLGLNLSGTGAEAGAPCGVAPQAR
ITGSSSAVAGQWPMQVSYIEGVHGGSGIVSOWILSAHCPSPSHHKEAYEVKGA
HQDYSYSEDAKYSTLNDIIPHPYLOBSGSDIALLOLRPIYFSXIRPICLPANA
SPFNGLHCTKTVGMHVAPSVLLTPKPLQLEVLISRETCLNVIDAPEHPHVQ
EDMVGAYVEGDACQDSGSGPLSCPVEBLMYLTGIVSGDAGCARNGVYTLASS
YASMIQSKTELOPRVVPQTOESQPSNLCGSHLAFSSAPQGLPIPLPLGALG
LISPMLESH"
1261..>1834
3'UTR
BASE COUNT 309 a 621 c 526 g 378 t
ORIGIN
Query Match 92.2%; Score 1537.4; DB 9; Length 1834;
Best Local Similarity 97.8%; Pred. No. 5,2e-299;
Matches 1625; Conservative 5; Mismatches 24; Indels 8; Gaps 7;
QY 1 CACACTGCTGCTTCGGATACTCCAGGCTCTCCGTTGCGGCGCTCCCTGCTTAGAG 60
Db CACACTGCTGCTTCGGATACTCCAGGCTCTCCGTTGCGGCGCTCCCTGCTTAGAG 81
QY 61 GCCAGCTTGGACACTTGTGCTCCCTTCCAGCCCGGATTTCTGGGATCTTCCCTCTGAG 120
Db GCCAGCTTGGACACTTGTGCTCCCTTCCAGCCCGGATTTCTGGGATCTTCCCTCTGAG 141
QY 82 GCCAGCTTGGACACTTGTGCTCCCTTCCAGCCCGGATTTCTGGGATCTTCCCTCTGAG 141
Db CCAACATCTGGGCTCTGCTTCGACACACCCCAAGGCTTCTACCTTGGCTGCTGAG 180
QY 142 CCAACATCTGGGCTCTGCTTCGACACACCCCAAGGCTTCTACCTTGGCTGCTGAG 201
Db 181 TCTGCCCAAGGGGCTTGTCTGCGGCTATGCGCMAGAGGGGCTCTGCGGCTTGGCA 240
QY 202 TCTGCCCAAGGGGCTTGTCTGCGGCTATGCGCMAGAGGGGCTCTGCGGCTTGGCA 260
Db 241 GCTGGGGGCTGTGGCCATTTCTGCTCTATCTGGAATTACTCCGCTGGGGGACAGAGGGA 300
QY 261 GCTGGGGGCTGTGGCCATTTCTGCTCTATCTGGAATTACTCCGCTGGGGGACAGAGGGA 320
Db 301 AGGGGCAAGAGCTCTGCGGCTGTGGCCCAAGCAAGCATCACAGTGGCAGCAGTGC 360
QY 321 AGGGGCAAGAGCTCTGCGGCTGTGGCCCAAGCAAGCATCACAGTGGCAGCAGTGC 380
Db 361 AGTGCCTGCTAGTGGCTTGTGGAGTCAAGCATCACTATGAAGGGCTCATGTGTGG 420
QY 381 AGTGCCTGCTAGTGGCTTGTGGAGTCAAGCATCACTATGAAGGGCTCATGTGTGG 440
Db 421 TGGCTCTCTGCTGTGAGGAGTGGGCTGTCAAGCTGTCACTGCTTCCCAAGGCA 480
QY 441 TGGCTCTCTGCTGTGAGGAGTGGGCTGTCAAGCTGTCACTGCTTCCCAAGGCA 500
Db 481 CCAAGAGAGCTATGAGTCAAGTGGGGGCCCACAGCTAGACTCTTCTCCGAGA 540
QY 501 CCAAGAGAGCTATGAGTCAAGTGGGGGCCCACAGCTAGACTCTTCTCCGAGA 560
Db 541 CGCCAGGTCAGCACTTGAAGACATCATCCCCACCCAGCTACTCTCGAGAGGGCTC 600
QY 561 CGCCAGGTCAGCACTTGAAGACATCATCCCCACCCAGCTACTCTCGAGAGGGCTC 620
Db 601 CCAAGGTCAGCTTCTCTCCACTGAGAGACCATCATCTTCTCCGCTCATTCG 660
QY 621 CCAAGGTCAGCTTCTCTCCACTGAGAGACCATCATCTTCTCCGCTCATTCG 680
Db 661 GCCCATCTGCTCTCTGAGGCAAGGCTCTTCCCAAGGCTCTCACTGCTGCTAC 720
QY 681 GCCCATCTGCTCTCTGAGGCAAGGCTCTTCCCAAGGCTCTCACTGCTGCTAC 740
Db 721 TGGCTGGGGTCAATGAGGCTCTGAGGCTCTGAGGCTCTGAGGCTCTGAGGCT 780
QY 741 TGGCTGGGGTCAATGAGGCTCTGAGGCTCTGAGGCTCTGAGGCTCTGAGGCT 800

```

```

QY 781 CGAGGCTCTGATCAGTGTGAGACGTGTAACTGCTGTATCAACATCGACGCCAAC 840
Db CGAGGCTCTGATCAGTGTGAGACGTGTAACTGCTGTATCAACATCGACGCCAAC 859
QY 841 CTGAGAGGCTGCACTTTGTCCAAAGAGACATGTGTGTCTGTGCTATGTGAGGGGGCA 900
Db CTGAGAGGCTGCACTTTGTCCAAAGAGACATGTGTGTGTGTGTATGTGAGGGGGCA 919
QY 901 AGAGAGCTGCAAGGCTGCTGAGGGGGCCACTGCTCCGCTGTGAGAGGCTCTGTGT 960
Db AGAGAGCTGCAAGGCTGCTGAGGGGGCCACTGCTCCGCTGTGAGAGGCTCTGTGT 979
QY 961 ACTGACGGGCAATTGAGACTGTGGAGATGCTGTGTGGGCCCCGCAACAGGCTGTGTGT 1020
Db ACTGACGGGCAATTGAGACTGTGGAGATGCTGTGTGGGCCCCGCAACAGGCTGTGTGT 1039
QY 1021 ACACTGTGCTTCCAGCTATGCTCTCTGTGATCCAAAGCAAGGTGACAACTCCAGCTTC 1080
Db ACACTGTGCTTCCAGCTATGCTCTCTGTGATCCAAAGCAAGGTGACAACTCCAGCTTC 1099
QY 1081 GTGTGTGCTCCCAAGCCAGAGGTCCAGCCGAGCAACCTGTGTGAGCAACCTGAG 1140
Db GTGTGTGCTCCCAAGCCAGAGGTCCAGCCGAGCAACCTGTGTGAGCAACCTGAG 1159
QY 1141 CTTTACGCTCTGCCCCAGGCCAGGGCTTGTGAGGCCCACTCTTCTGCTCTGAGGCC 1200
Db CTTTACGCTCTGCCCCAGGCCAGGGCTTGTGAGGCCCACTCTTCTGCTCTGAGGCC 1219
QY 1201 TGGCTGTGGGCTCTCTCTCCCAATGCTGAGAGCACTGAGTGGCCCTACTTCCAGGA 1260
Db TGGCTGTGGGCTCTCTCTCCCAATGCTGAGAGCACTGAGTGGCCCTACTTCCAGGA 1279
QY 1261 TGGATCATCACTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1320
Db TGGATCATCACTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1339
QY 1280 TGGATCATCACTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1380
Db 1321 GCCTGACTTGTGAGCACTCTCTTCCATGAGACTGTGAGAGAGAGAGAGAGAGAGAG 1399
QY 1340 GCCTGACTTGTGAGCACTCTCTTCCATGAGACTGTGAGAGAGAGAGAGAGAGAGAG 1400
Db 1381 CTTTGAAGCCCATCTTCTGAGGTGTGCTTTTGGAGACATCACTGAGAGAGAGAGAG 1440
QY 1400 CTTTGAAGCCCATCTTCTGAGGTGTGCTTTTGGAGACATCACTGAGAGAGAGAGAG 1459
Db 1441 CTGCTGTGAGCACTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1500
QY 1460 CTGCTGTGAGCACTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1519
Db 1501 CGAGGCTCTGAGGAG--TCTGTGAGACTTGTGTATGAAGAGAGAGAGAGAGAGAG 1558
QY 1520 CGAGGCTCTGAGGAGCTCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1578
Db 1559 GTTTCCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1618
QY 1579 GTTTCCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1635
Db 1619 CTTTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1660
QY 1636 CTTTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1677
Db 1636 CTTTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG

```

```

RESULT 8
BC001462 1809 bp mRNA linear PRI 12-JUL-2001
LOCUS
DEFINITION
Homo sapiens, protease, serine, 8 (prolactin), clone MGC:2133
IMAGE:313832, mRNA, complete cds.
ACCESSION
BC001462
VERSION
BC001462.1 GI:12655206
KEYWORDS
MGC.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```



REFERENCE  
AUTHORS  
TITLE  
JOURNAL

Mammalia: Eutheria: Primates: Catarrhini, Homiidae: Homo.  
1 (bases 1 to 1809)  
Strausberg, R.  
Direct Submission  
Submitted (12-DEC-2000) National Institutes of Health, Mammalian  
Gene Collection (MGC), Cancer Genomics Office, National Cancer  
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
USA

REMARK  
COMMENT

NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
Contact: MGC help desk  
Email: [cgabs-remail.nih.gov](mailto:cgabs-remail.nih.gov)  
Tissue Procurement: ATCC  
CDNA Library Preparation: Rubin Laboratory  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILNI)  
DNA Sequencing by: National Institutes of Health Intramural  
Sequencing Center (NISC),  
Gaithersburg, Maryland:  
Web site: <http://www.nisc.nih.gov/>  
Contact: [nisc\\_mgc@nsl.nih.gov](mailto:nisc_mgc@nsl.nih.gov)  
Shevchenko, Y., Weisberg, K.D., Beckstrom-Sternberg, S.M.,  
Benjamin, B., Blakesley, R.W., Bouffard, G.G., Brinkley, C., Brooks, S.,  
Dietrich, N.L., Guan, X., Gupta, J., Ho, S.-L., Karlins, E., Legaspi, R.,  
Lim, M., Maduro, Q.L., Masello, C., Maerrian, S.D., McCloskey, J.C.,  
McDowell, J., Pearson, R., Snyder, B., Stantipop, S., Thomas, P.J.,  
Tiongson, E.E., Touchman, J.W., Tsurgoun, C., Vogt, J.D., Walker, M.A.,  
Zhang, L.-H. and Green, E.D.

Clone distribution: MGC clone distribution information can be found  
through the I.M.A.G.E. Consortium/ILNI at: <http://image.llnl.gov>  
Series: IRAL Plate: 4 Row: 3 Column: 3  
This clone was selected for full length sequencing because it  
passed the following selection criteria: matched mRNA gi: 1143193.  
Location/Qualifiers

FEATURES  
source

1. 1809  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="LocustID:5652"  
/db\_xref="taxon:9606"  
/clone="MGC:2133 IMAGE:3138532"  
/tissue\_type="placenta, choriochorionoma"  
/clone\_id="NIH\_MGC\_21"  
/lab\_host="DH10B-R"  
/note="Vector: pOT7"  
185..1216  
/codon\_start=1  
/product="protease, serine, 8 (protrasein)"  
/protein\_id="AAH01462.1"  
/db\_xref="GI:12655207"  
/translation="MAQKGLDPLGQAGAVALLYLGLNSGTAGAEAPCGVAPQAR  
ITGSSAVAGQMPWQVSYIEGVHVGSGSLVSEOWYLSAHPFSPSHHAEVAVKGA  
HOLDYSYEDAKYSLKDIIPHSYLOEGSDIALLQSLRPTTFPSYRIPILPANA  
SPNGLHCTVGTGNGVAPSVSLTPKLOQLVPLISPTCNCLYNIDKPREPHVQ  
EIMVCAGYEGGKACQGDSDGRLSPVGLWYLTGIVSWBACGARNRPGVYTIASS  
YASWIOSKVTLEQPRVPTQESQPSNLCGSHLAFSAAPAGLLRPLIFLGLALG  
LISPMLESH"

## CDS

BASE COUNT 322 a 605 c 517 g 365 t  
ORIGIN

Query Match 90.9%; Score 1516.4; DB 9; Length 1809;  
Best Local Similarity 97.8%; Pred. No. 8.8e-295;  
Matches 1593; Conservative 5; Mismatches 24; Indels 7; Gaps 6;

QY 34 CCGTTCGCGCGCTCCCTGCTAGAGGCGAGCTTGAACATCTGCTGCCCTTTCCAGC 93  
DB 10 CCGTTCGCGCGCTCCCTGCTAGAGGCGAGCTTGAACATCTGCTGCCCTTTCCAGC 69  
QY 94 CCGATTCGGAGATCTTCTCTGAGCCAAACATCTGGATCTGCTTGAACACCC 153  
DB 70 CCGGATTCGGAGATCTTCTCTGAGCCAAACATCTGGATCTGCTTGAACACCC 129  
QY 154 AAGGCTTCTACTTCCGCTGAGATCTGCGCCGAGGCGCTTGTCTGGGCGCATGAC 213  
DB 130 AAGGCTTCTACTTCCGCTGAGATCTGCGCCGAGGCGCTTGTCTGGGCGCATGAC 189

QY 214 CMAAAGGGGGCTCTGGGGCTGGGAGCTGGGGGCTGTGGCCATCTGCTATCTTGG 273  
DB 190 CCAGAAAGGGGGCTCTGGGGCTGGGAGCTGGGGGCTGTGGCCATCTGCTATCTTGG 249  
QY 274 ATTATCTCCGCTGGGGAGCAGAGCGGAGAGGCTTCCTGGGTGGGGCCCCCA 333  
DB 250 ATTACTCCGCTGGGGAGCAGAGCGGAGAGGCTTCCTGGGTGGGGCCCCCA 309  
QY 334 AGACGATCAGAGGTGGAGAGAGTCCCGCTGAGTGGCCCTGGCAGTCCAGT 393  
DB 310 AGACGATCAGAGGTGGAGAGAGTCCCGCTGAGTGGCCCTGGCAGTCCAGT 369  
QY 394 CACCTATGAGGCGTCCATGTGTGTGGTCTCTCGGTCTGAGCAGTGGGTGTGC 453  
DB 370 CACCTATGAGGCGTCCATGTGTGTGGTCTCTCGGTCTGAGCAGTGGGTGTGC 429  
QY 454 AGCTGCTCACTCTTCCAGGAGCAGCAGAGAGGCTTATGAGTCAAGCTGGGGC 513  
DB 430 AGCTGCTCACTCTTCCAGGAGCAGCAGAGAGGCTTATGAGTCAAGCTGGGGC 489  
QY 514 CCACGAGTAACTCTTCTTCCAGAGAGCCCAAGTCAAGCCTTGAAGAGATATCC 573  
DB 490 CCACGAGTAACTCTTCTTCCAGAGAGCCCAAGTCAAGCCTTGAAGAGATATCC 549  
QY 574 CCACCCAGCTACTCCAGAGAGGCTCCAGGGGCGCATTGACCTTCCATCTAGCAG 633  
DB 550 CCACCCAGCTACTCCAGAGAGGCTCCAGGGGCGCATTGACCTTCCATCTAGCAG 609  
QY 634 ACCCATCACTTCTCCGCTACATCCGGCCATCTGCTCCCTGAGCCAGCCCTCTT 693  
DB 610 ACCCATCACTTCTCCGCTACATCCGGCCATCTGCTCCCTGAGCCAGCCCTCTT 669  
QY 694 CCCCAGGCGCTCCACTGACTGTCACTGAGTGGGTATGTGGCCCCCTCATGAGCT 753  
DB 670 CCCCAGGCGCTCCACTGACTGTCACTGAGTGGGTATGTGGCCCCCTCATGAGCT 729  
QY 754 CCGTACGCGCCAGGCACTGAGAGAGTGGCTGTGATGAGTGGAGAGCGGTGA 813  
DB 730 CCGTACGCGCCAGGCACTGAGAGAGTGGCTGTGATGAGTGGAGAGCGGTGA 788  
QY 814 ACTGCTGTACAACTGACGCGCAAGCCTGAGAGCCGCACTTGTCCAAAGAGCATGG 873  
DB 789 ACTGCTGTACAACTGACGCGCAAGCCTGAGAGCCGCACTTGTCCAAAGAGCATGG 848  
QY 874 TGTGTGCTGTATGTGAGAGGGGGGCAAGAGCCTTCCAGGGTGACTTGGGGGCCAC 933  
DB 849 TGTGTGCTGTATGTGAGAGGGGGGCAAGAGCCTTCCAGGGTGACTTGGGGGCCAC 908  
QY 934 TCTCGGCGCTGTGAGAGGGTCTGTGGTCTCTGAGCGGGCATTTGTAGCTGGGGAGATGGCT 993  
DB 908 TCTCGGCGCTGTGAGAGGGTCTGTGGTCTCTGAGCGGGCATTTGTAGCTGGGGAGATGGCT 968  
QY 994 GTGGGGCCGCAAGAGGCTGTGTGTACATCTTGGCTTCCAGTATGTGCTTGGATCC 1053  
DB 969 GTGGGGCCGCAAGAGGCTGTGTGTACATCTTGGCTTCCAGTATGTGCTTGGATCC 1028  
QY 1054 AAAGCAAGTGCAGAACTCCAGCTCTGTGTGTCGCCCAAAACCAGAGTCCAGCCCG 1113  
DB 1029 AAAGCAAGTGCAGAACTCCAGCTCTGTGTGTCGCCCAAAACCAGAGTCCAGCCCG 1088  
QY 1114 ACAGCACTGTGAGGAGGCACTGAGGCTTCACTGCTGAGCCAGGCGCTTGTGTA 1173  
DB 1088 ACAGCACTGTGAGGAGGCACTGAGGCTTCACTGCTGAGCCAGGCGCTTGTGTA 1148  
QY 1174 GGGCCATCTTCTTCTGCTCTGAGGCTGTGGGCTCTGAGGCTCTCCATGGCTCAGCG 1233  
DB 1149 GGGCCATCTTCTTCTGCTCTGAGGCTGTGGGCTCTGAGGCTCTCCATGGCTCAGCG 1208  
QY 1234 AGCATGAGCTGGGCTTCTTCCAGAGTGAATGATCATCACTCAAGAGCAGAGGCTGTG 1293  
DB 1209 AGCATGAGCTGGGCTTCTTCCAGAGTGAATGATCATCACTCAAGAGCAGAGGCTGTG 1268

QY	1291	CTTCCCGMAGGCGCTTTGAGCCGAGGGCCCTGACTGAGGACATCTCTTCAGAGACTC	1353
Db	1269	CTTTCCTCGATGGCGCTTTGAGCCGAGGGCCCTGACTGAGGACATCTCTTCAGAGACTC	1328
QY	1354	TGCGGGAGGCGTGGGGCCCCCATCTTGATCTTTGAGCCCATCTCTCGGGTGTGCTTTTGG	1413
Db	1329	TGCGGGAGGCGTGGGGCCCCCATCTTGATCTTTGAGCCCATCTCTCGGGTGTGCTTTTGG	1388
QY	1414	GACCATCATCTGAGAGTCAGAGATTTTACTGCTCTGTAGCAATGGCCAGAGCCTCTGGCCCC	1473
Db	1389	GACCATCATCTGAGAGTCAGAGATTTTACTGCTCTGTAGCAATGGCCAGAGCCTCTGGCCCC	1448
QY	1474	TCAMCCACCAATGAGCCAGGCCCATTTGSGGAGANTCCCTGGGGAG--TCCGGGAGCCTTGGYT	1531
Db	1449	TCACCCACCAATGAGCCAGGCCCATTTGSGGAGANTCCCTGGGGAGCTCTGGAGACCTTGGCT	1508
QY	1532	ATGAAAAATGAGACCCCTGGGTTCCCACTGTTTCTGMAAACACTGCTTCCCGGCGCGCTTCC	1591
Db	1509	ATGAAAAATGAGACCCCT--GGCTCCCACTGTGTTCTGMAAACACTGCG--TCCGGGCGCGCTTCC	1566
QY	1592	CAGACTMAGTAGAGCAATTTTTTTTGCNNTTCCCTGTGTTTTGGGTGGCAACTTTT	1651
Db	1567	CAGACT--GATGAGGACA--TCTCTCGCCCTCTCCTCTGTGTTCTGGGCTGGGGCGACACTTT	1624
QY	1652	TGGAAGTTT 1660	
Db	1625	GTGCAGCTT 1633	

Qy	313	TYCCTGGGtGTGGCCCCCAAGCAAGCATCACAGGtGGAGGAGTGCAGTCCGGGTCA	372
	256	TCCTGGGtGTGGCCCCCAAGCAAGCATCACAGGtGGAGGAGTGCAGTCCGGGTCA	315
Qy	373	GTGGCCCTGGCAGGTGCAGTCACTATGAAAGGCGTTCATGTGTGTGGtCTCTCGT	432
	316	GTGGCCCTGGCAGGTGCAGTCACTATGAAAGGCGTTCATGTGTGTGGtCTCTCGT	375
Qy	433	GTCTGAGCAGTGGGTGTCTGACGTCTCACTGTCTTCCCAAGGAGTCCACCAAGAAAGC	492
Db	376	GTCTGAGCAGTGGGTGTCTGACGTCTCACTGTCTTCCCAAGGAGTCCACCAAGAAAGC	435
Qy	493	CTATGAGGTAAAGTGGGGGCCCAACAGCTAGA CTCTACTCCGAGGAGCCAAAGTCAAG	552
Db	436	CTATGAGGTAAAGTGGGGGCCCAACAGCTAGA CTCTACTCCGAGGAGCCAAAGTCAAG	495
Qy	553	CACCTTAAGGAGCATCTATCCCAACCCCAAGCTACTTCAGAGAGGGCTCCAGGGCGACAT	612
Db	496	CACCTTAAGGAGCATCTATCCCAACCCCAAGCTACTTCAGAGAGGGCTCCAGGGCGACAT	555
Qy	613	TGCACCTCTTCCAACTCAGCAGACCCATCACCTTCTCCGCTACATCCGGCCCATCTGCTT	672
Db	556	TGCACCTCTTCCAACTCAGCAGACCCATCACCTTCTCCGCTACATCCGGCCCATCTGCTT	615
Qy	673	CCCTGAGGCAAGGCGCTCTCCCAAGGGCGTCACTGACGTCTCACTGGGTGGGGTCA	732
	616	CCCTGAGGCAAGGCGCTCTCCCAAGGGCGTCACTGACGTCTCACTGGGTGGGGTCA	675

[illegible]

Oy	733	TGTGGCCCTCAGTGAAGCTCTCGAAGCCCAAGCCACTCAGCAATCTGAGGTGCTCT	792
Db	676	TGTGGCCCTCAGTGAAGCTCTCGAAGCCCAAGCCACTCAGCAATCTGAGGTGCTCT	735
Oy	793	GATACATCTGAGACCTGTGTATCTGCTGTATCAACATCTGACGCCAABCCCTGAGGAGCCG	852
Db	736	GATACATCTGAGACCTGTGTATCTGCTGTATCAACATCTGACGCCAABCCCTGAGGAGCCG	794
Oy	853	ACTTTGTCCAAAGAGCATGGTGTGTGCTGAGCTATGTGAGGGGGGCAAGAGCCGCTGCC	912
Db	795	ACTTTGTCCAAAGAGCATGGTGTGTGCTGAGCTATGTGAGGGGGGCAAGAGCCGCTGCC	854
Oy	913	AGGGTGACTCTGGGGGGCCCACTCTCTGCTGCTGTGTGAGGGGTCTTGGTATCTGACGGGCA	972
Db	855	AGGGTGACTCTGGGGGGCCCACTCTCTGCTGCTGTGTGAGGGGTCTTGGTATCTGACGGGCA	914
Oy	973	TTGTGACTGGGGAGATGCTGTGTGGGGCCGCCAACAGGCTGTGTATCACTCTGGGCT	1032
Db	915	TTGTGACTGGGGAGATGCTGTGTGGGGCCGCCAACAGGCTGTGTATCACTCTGGGCT	974
Oy	1033	CCAGCTATGCTCTCTGGAATCCAAAGCAAGTGAACAGAACTTCACGCTCTGTGGTGGCCCC	1092
Db	975	CCAGCTATGCTCTCTGGAATCCAAAGCAAGTGAACAGAACTTCACGCTCTGTGGTGGCCCC	1034
Oy	1093	AAACCCAGAGATCCACAGCCCGACAGCAACCTCTGTGACAGCCACACTGGCCCTTCAGCTCTG	1152
Db	1035	AAACCCAGAGATCCACAGCCCGACAGCAACCTCTGTGACAGCCACACTGGCCCTTCAGCTCTG	1094
Oy	1153	CCCCAGCCCAAGGCTTGTGTAGGCCCATCTTTTCTCTGCTCTGTGGGCTTGGGCTCTGGGCC	1212
Db	1095	CCCCAGCCCAAGGCTTGTGTAGGCCCATCTTTTCTCTGCTCTGTGGGCTTGGGCTCTGGGCC	1154
Oy	1213	TCTCTCCCCCATGAGCTCAGCGAGCAGAGCTGAGCCCTACTTCTCAGAGTGAATCATATAC	1272
Db	1155	TCTCTCCCCCATGAGCTCAGCGAGCAGAGCTGAGCCCTACTTCTCAGAGTGAATCATATAC	1214
Oy	1273	ACTCAAGAGCAGAGCTGTGTCTTTCCTGTAGGCTCTTTGGACCCAGGGCTGACTTGAG	1332
Db	1215	ACTCAAGAGCAGAGCTGTGTCTTTCCTGTAGGCTCTTTGGACCCAGGGCTGACTTGAG	1274
Oy	1333	CCACTCCCTCTTCAAGACTCTGCGGGAGGCTGGGGCCCATCTTGATCTTTGAGCCCAT	1392
Db	1275	CCACTCCCTCTTCAAGACTCTGCGGGAGGCTGGGGCCCATCTTGATCTTTGAGCCCAT	1334

```

Qy 1393 TCTTCTGGGTGCTCTTTTGGGACCATCATCTGAGAGTCCAGGTTTACTGCTGTAGCA 1452
Db 1335 TCTTCTGGGTGCTCTTTTGGGACCATCATCTGAGAGTCCAGGTTTACTGCTGTAGCA 1394
Qy 1453 ATGGCCAGAGCCTCTGGGCCCCCTGACCCCACTGAGACAGCCCATTTGSSCCAGNATCCGGG 1512
Db 1395 ATGGCCAGAGCCTCTGGGCCCCCTGACCCCACTGAGACAGCCCATTTGCGGAGCTCCGGG 1454
Qy 1513 GAG--TCTCTGGGACCTTGGTATGAAAATGAGCCCTGGGTTCCACCTGTTTCTNGAGA 1570
Db 1455 GAGTCTCTGGGACCTCTGGGCCCCCTGACCCCACTGAGACAGCCCATTTGCGGAGG 1513
Qy 1571 CTGCTTCCCGGCCCCCTTCCCAAGCTNGATGAGACATTTTTCCTCCCTTCCCTGTG 1630
Db 1514 CTGC-TCCCGGCCCCCTTCCCAAGCT-GATGAGACACA-TCTCTCTGCTCCCTCTCTGTG 1570
Qy 1631 TTTTGGGTTGGGCACTTTTGGAGTTT 1660
Db 1571 TTTTGGGTTGGGCACTTTTGGAGTTT 1600

RESULT 10
AX675579 1726 bp DNA linear PAT 27-MAR-2003
LOCUS Sequence 29 from Patent WO02055704.
ACCESSION AX675579
VERSION AX675579.1 GI:29333566
KEYWORDS
SOURCE
ORGANISM Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 Padigaru, M., Li, L., Zerhusen, B. D., Casman, S. J., Shenoy, S.,
Spreck, K. A., Zhong, M., Gangoli, E. A., Burgess, C. E., Patturajan, M.,
Vernet, C. A., Taylor, S., Tenhove, V. T., Miller, C. E., Guo, X.,
Bolog, F. L., Grose, W. M., Alsobrook, J. P., Gerlach, V.,
Edingermark, S., Rothenberg, M. E., Ellerman, K., Macdougall, J.,
Maljaner, U., Millet, I., Peyman, J., Smithson, G., Gunther, E. and
Stone, D. J.
Proteins, polynucleotides encoding them and methods of using the
same
JOURNAL Patent: WO 02055704-A 29 18-JUL-2002;
Curagen Corporation (US)
FEATURES
source 1..1726
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
BASE COUNT 284 a 581 c 499 g 362 t
ORIGIN
Query Match 78.2%; Score 1305; DB 6; Length 1726;
Best Local Similarity 91.0%; Pred. No. 2,9e-252;
Matches 1513; Conservative 5; Mismatches 28; Indels 116; Gaps 8;

```

```

Db 261 GCTGGGGCTGTGGCATTCTGCTATCTTGATTAATCCGGTCGGGAGCAGAGCGGA 320
Qy 301 AGGGGAGAAAGCTCTCTGGGTGTGGCCCCCAGAGACGATACAGGTGGAGCAGTGC 360
Db 321 AGGGGAGAAAGCTCTCTGGGTGTGGCCCCCAGAGACGATACAGGTGGAGCAGTGC 380
Qy 361 AGTGCCTGCTGATGAGTCCCTGGAGGTGAGCATCACTTAAGAGGCTCATGTGTGG 420
Db 381 AGTGCCTGCTGATGAGTCCCTGGAGGTGAGCATCACTTAAGAGGCTCATGTGTGG 440
Qy 421 TGGCTCTCTGATGAGCAGTGGTGTGATGCTGCTCACTGCTTCCAGCAGAGA 480
Db 441 TGGCTCTCTGATGAGCAGTGGTGTGATGCTGCTCACTGCTTCCAGCAGAGA 500
Qy 481 CCACAAAGAAAGCTATGAGGTCAAGTGGGGGCCACAGCTAAGTCTTCCAGAGA 540
Db 501 CCACA----- 505
Qy 541 CGCAGAGTCAAGACCCCTGAAGGACATCATCCCCACCCAGCTACCTCAGAGGGCTC 600
Db 506 -----AGGCTC 512
Qy 601 CAGGGGCAATTTGCACTCTCCAACTCAGACAGACCCATCACTTCTCCGCTACATCCG 660
Db 513 CAGGGGCAATTTGCACTCTCCAACTCAGACAGACCCATCACTTCTCCGCTACATCCG 572
Qy 661 GCCCATCTGCTCTCTCTGAGCCCAAGCCCTCTTCCCAAGGCTCTCACTGATGAC 720
Db 573 GCCCATCTGCTCTCTCTGAGCCCAAGCCCTCTTCCCAAGGCTCTCACTGATGAC 632
Qy 721 TGGCTGGGCTCAATGAGGCCCCCTCAGTGAAGCTCTGAGCCCAAGGCTCAGCAACT 780
Db 633 TGGCTGGGCTCAATGAGGCCCCCTCAGTGAAGCTCTGAGCCCAAGGCTCAGCAACT 692
Qy 781 CGAGTGCTCTGATCAGTGTGAGACGTGTACTGCTGTACAACTCAGCAGCAGC 840
Db 693 CGAGTGCTCTGATCAGTGTGAGACGTGTACTGCTGTACAACTCAGCAGCAGC 751
Qy 841 CTGAGAGGCGGACATTTGTCCAAAGAGACATGTGTGTGTGTGTGTGTGTGTGTGT 900
Db 752 CTGAGAGGCGGACATTTGTCCAAAGAGACATGTGTGTGTGTGTGTGTGTGTGTGT 811
Qy 901 AGGAGCGCTGCGAGGGGTGCTCTGGGGGCCACTCTCCGCTCTGAGAGGTCTCTGT 960
Db 812 AGGAGCGCTGCGAGGGGTGCTCTGGGGGCCACTCTCTGCTGTGAGAGGTCTCTGT 871
Qy 961 AACTGACGGGCAATTTGTGAGCTGGGAGATGCTGTGGGGCCCGCAACAGCCTGTGTGT 1020
Db 872 AACTGACGGGCAATTTGTGAGCTGGGAGATGCTGTGGGGCCCGCAACAGCCTGTGTGT 931
Qy 1021 AACTTGTGCTCTGAGCTATGCTCTCTGATCAAAAGAGTGAAGAGTCAAGCTTC 1080
Db 932 AACTTGTGCTCTGAGCTATGCTCTCTGATCAAAAGAGTGAAGAGTCAAGCTTC 991
Qy 1081 GTGTGATGCCCAAAACCAAGAGTCCAGCCCAAGAGAACCTGTGTGGAGCCACTGG 1140
Db 992 GTGTGATGCCCAAAACCAAGAGTCCAGCCCAAGAGAACCTGTGTGGAGCCACTGG 1051
Qy 1141 CTTTGAAGTCTGCCCCAGCCCAAGGCTTGTGAGGCCATCTTTTCTGCTCTGGGCC 1200
Db 1052 CTTTGAAGTCTGCCCCAGCCCAAGGCTTGTGAGGCCATCTTTTCTGCTCTGGGCC 1111
Qy 1201 TGGCTCTGGGCTCTCTTCTCCCAATGCTCAAGAGACATGAGTGGCTTCACTTCAAGA 1260
Db 1112 TGGCTCTGGGCTCTCTTCTCCCAATGCTCAAGAGACATGAGTGGCTTCACTTCAAGA 1171
Qy 1261 TGGATGATCACTCAAGAGCAGAGGCTGTCTCTTCTGATGAGTGGCTTCAAGCCAGG 1320
Db 1172 TGGATGATCACTCAAGAGCAGAGGCTGTCTCTTCTGATGAGTGGCTTCAAGCCAGG 1231
Qy 1321 GCTTGAATTAAGCACTCTTCTTCAAGACTTGTGCGGAGGCTGGGGCCCATCTTGTAT 1380

```

Db 1232 GCCTGACTTGAGCCACTCTTCTTCAGAGACTGCGGAGGCTGGGCCCATCTTGAT 1291

QY 1381 CTTTGAGCCCATCTTCTGGAGTGTCTTTTGGAGACCATGACTGAGAGTGTGA 1440

Db 1292 CTTTGAGCCCATCTTCTGGAGTGTCTTTTGGAGACCATGACTGAGAGTGTGA 1551

QY 1441 CTCCTGTAGCAATGGCCAGAGCCTCTGAGCCCTCAGACCAATGAGCAGCCCATTTGS 1500

Db 1352 CTCCTGTAGCAATGGCCAGAGCCTCTGAGCCCTCAGACCAATGAGCAGCCCATTTGC 1411

QY 1501 CGAGNTCTGGGAG--TCTGGAGACCTTGATGATGAATAAGACCTTGATTTCCACT 1558

Db 1412 CGAGCTCTGGGAGCTCTGGAGACCTTGATGATGAATAAGACCTTGATTTCCACT 1470

QY 1559 GTTCTGAGAGACTGCTTCCGAGCCCTTCCAGACTGATGATGACATTTTGTG 1618

Db 1471 GTTCTGAGAGACTGCTTCCGAGCCCTTCCAGACTGATGATGACATTTTGTG 1527

QY 1619 CTTTCTGCTGTGTTTGGGTTGGGCAACTTTTGAAGTTT 1660

Db 1528 CTTCTCCTGTGTTTGGGTTGGGCAACTTTTGTGAGCTT 1569

RESULT 11  
AC009088/c 127769 bp DNA linear PRI 29-MAR-2003  
LOCUS Homo sapiens chromosome 16 clone RP11-388M20, complete sequence.  
DEFINITION AC009088  
AC009088  
AC009088.9 GI:29366934  
VERSION HTG.  
KEYWORDS Homo sapiens (human)  
SOURCE Homo sapiens  
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE  
AUTHORS 1 (bases 1 to 127769)  
DOE Joint Genome Institute, Stanford Human Genome Center and Los Alamos National Laboratory.  
TITLE Direct Submission  
JOURNAL Unpublished  
AUTHORS 2 (bases 1 to 127769)  
DOE Joint Genome Institute.  
REFERENCE  
AUTHORS Direct Submission  
JOURNAL Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
3 (bases 1 to 127769)  
DOE Joint Genome Institute.  
REFERENCE  
AUTHORS Direct Submission  
JOURNAL Submitted (03-APR-2002) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
4 (bases 1 to 127769)  
DOE Joint Genome Institute, Stanford Human Genome Center and Los Alamos National Laboratory.  
REFERENCE  
AUTHORS Direct Submission  
JOURNAL Submitted (18-MAR-2003) DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
5 (bases 1 to 127769)  
DOE Joint Genome Institute, Stanford Human Genome Center and Los Alamos National Laboratory.  
TITLE Direct Submission  
JOURNAL Submitted (29-MAR-2003) DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
On Mar 29, 2003 this sequence version replaced gi:29029216.  
Draft Sequence Produced by DOE Joint Genome Institute  
www.jgi.doe.gov  
Finishing Completed at Stanford Human Genome Center and Los Alamos National Laboratory  
www.snhg.stanford.edu  
Quality: Phrap Quality >=40.99.7% of Sequence;  
Estimated Total Number of Errors is 0.21  
NOTE: This insert is not the entire sequence of the clone (entire sequence is 233.4kb). It is clipped at the overlap with AC135050.  
The number of bases overlapped is 2575.  
Location/Qualifiers

source 1. 127769  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="16"  
/clone="RP11-388M20"

BASE COUNT 30235 a 31458 c 34056 g 32020 t

ORIGIN

Query Match 52.8%; Score 880.2; DB 9; Length 127769;  
Best Local Similarity 84.5%; Pred. No. 3.4e-167;  
Matches 1157; Conservative 3; Mismatches 24; Indels 186; Gaps 8;

QY 472 CAGCGAGCAACCAAGAAAGCTATGAGTCAAGCTGGGGGCCCAAGCTAGACTCTTA 531

Db 15871 CAGCGAGCAACCAAGAAAGCTATGAGTCAAGCTGGGGGCCCAAGCTAGACTCTTA 15812

QY 532 CTCGAGAGAGCCCAAGAGTCAAGCCTTGAAAGACATCATCCCAACCCAGCTACTCCA 591

Db 15811 CTCGAGAGAGCCCAAGAGTCAAGCCTTGAAAGACATCATCCCAACCCAGCTACTCCA 15752

QY 592 GAGGGGCTCCAGGGGAGCAATTGCACTCCCAACTCAGAGAGCCATCACTTCTCCG 651

Db 15751 GAGGGGCTCCAGGGGAGCAATTGCACTCCCAACTCAGAGAGCCATCACTTCTCCG 15692

QY 652 CTACATCCGAGCCCATCTGCTCCCTGACAGCAAGCCTCTTCCCAAGGCTCCACTG 711

Db 15691 CTACATCCGAGCCCATCTGCTCCCTGACAGCAAGCCTCTTCCCAAGGCTCCACTG 15632

QY 712 CACTGTCACTGGCTGGGGTCAATGTGCCCCCT----- 743

Db 15631 CACTGTCACTGGCTGGGGTCAATGTGCCCCCTCAGGTGAGTGGAGAGCTGGGTGCTAG 15572

QY 744 -----CA 745

Db 15571 AGGTGAGAGGGGACCTGACTTGGGAGAGGGCCAGGGAAGCTTTTACCCACA 15512

QY 746 GTGAGCCTCCTGAGCGCCCAAGACCATGAGCAACTGAGGTGCTGTGACTGCTGAG 805

Db 15511 GTGAGCCTCCTGAGCGCCCAAGACCATGAGCAACTGAGGTGCTGTGACTGCTGAG 15452

QY 806 AGGTGTAATCTGCTGTACCAATGCAAGCGCAAGCTTGAAGAGCGCACTTTGTCAAGA 865

Db 15451 AGGTGTAATCTGCTGTACCAATGCAAGCGCAAGCTTGAAGAGCGCACTTTGTCAAGA 15393

QY 866 GGAATGAGTGTGCTGAGTATGAGAGGGGAGCAAGAGCGCTG----- 911

Db 15392 GGAATGAGTGTGCTGAGTATGAGAGGGGAGCAAGAGCGCTGAGTAACACAGG 15333

QY 912 ----- 911

Db 15332 CCGGGGGGAGATGACAGTGAACCTGGGAAAGAGGCTGGCCCGTCTGATGGCT 15273

QY 912 -----CAGGTGACTTGGGGGCCCACTTCTGCTGCTGTGAAGG 952

Db 15272 GCTGTGAGCTTCTCTCTAGAGGTGACTCTGGGGGCCCACTTCTGCTGTGAAGG 15213

QY 953 TCTCTGTGATCAGACCGGCACTTGTGAGCTGGGGAGATGCTGAGGGCCGCAACAGGCC 1012

Db 15212 TCTCTGTGATCAGACCGGCACTTGTGAGCTGGGGAGATGCTGAGGGCCGCAACAGGCC 15153

QY 1013 TGGTGTGTAACACTGTGGCCTCAGACTATGCTCTGATTCAGAAAGAGGTGACAGAACT 1072

Db 15152 TGGTGTGTAACACTGTGGCCTCAGACTATGCTCTGATTCAGAAAGAGGTGACAGAACT 15093

QY 1073 CAGAGCTGTGTGTGCTGCCAAACCAAGAGTCCAGAGCCGACAGCAACTCTGTGGCAG 1132

Db 15092 CAGAGCTGTGTGTGCTGCCAAACCAAGAGTCCAGAGCCGACAGCAACTCTGTGGCAG 15033

QY 1133 CCACTGAGCTTGAAGCTCTGAGCCCAAGAGGCTTGTGAGGCGCACTTTTCTGCGC 1192

Db 15032 CCACTGAGCTTGAAGCTCTGAGCCCAAGAGGCTTGTGAGGCGCACTTTTCTGCGC 14973

OY	1193	TCGTGGGCTCGTACCTCGGGCCCTCCCTCCCATGGATCAGGAGACACTGAGCTGGGCCCTAC	1252
Db	14972	TCGTGGGCTCGTACCTCGGGCCCTCCCTCCCATGGATCAGGAGACACTGAGCTGGGCCCTAC	14913
OY	1253	TTCCAGATGAGATGCATCACACTCAAGACAGAGAGCTGGTCCCTCCATGAGCCCTTAC	1312
Db	14912	TTCCAGATGAGATGCATCACACTCAAGACAGAGAGCTGGTCCCTCCATGAGCCCTTAC	14853
OY	1313	GACCCAGGGCTGACCTTGAGACCACTCCCTTCCTTCAGAGCTCTGCGGGAGAGCTGGGGCCCC	1372
Db	14852	GACCCAGGGCTGACCTTGAGACCACTCCCTTCCTTCAGAGCTCTGCGGGAGAGCTGGGGCCCC	14793
OY	1373	ATCTTGATCTTTGAGCCCATCTCTCTGGAGTGCTCTTTTGGAGCACTACCTGAGAGTCAG	1432
Db	14792	ATCTTGATCTTTGAGCCCATCTCTCTGGAGTGCTCTTTTGGAGCACTACCTGAGAGTCAG	14733
OY	1433	GAGTTTAACTGCTCTGTAAGCAATGGCCAGAGCTCTGGCCCCCTCAGCCACCATGAGCCAGC	1492
Db	14732	GAGTTTAACTGCTCTGTAAGCAATGGCCAGAGCTCTGGCCCCCTCAGCCACCATGAGCCAGC	14673
OY	1493	CCATTGGSCGAGATTCCTCGGGGAG--TCCTGGGAGCTTGGTATGAAGAAATGAGCGCCGGGT	1550
Db	14672	CCATTGGSCGAGCTCCTCGGGGAGCTCCTGGGAGACCTTTGGCTATGAAGAAATGAGCGCCT--GGC	14614
OY	1551	TCCCACTGTGTTCTGGAAGACTGCTTCCCGGCCCGCCTTCCAGACTGATGAGCAATT	1610
Db	14613	TCCCACTGTGTTCTGGAAGACTGC--TCCCGGCCCGCCTGCGCCAGACT--GATGAGCA--T	14557
OY	1611	TTTTTGGCAATTCCTCGTGTCTTTTGGATGGGCACTTTTGAAGTTT	1660
Db	14556	CTCTCTGCGCTCTCCCTGATGTTCTGGGCTGGGGCCACCTTTGCGAGCTT	14507

RESULT 12				
LOCUS	HSU33446	7008 bp	DNA	linear
DEFINITION	Human prostaticin gene, complete cds.			
ACCESSION	U33446			
VERSION	U33446.1	GI:1143193		
KEYWORDS	.			
SOURCE	Homo sapiens (human)			
ORGANISM	Homo sapiens			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
AUTHORS	1 (eites)			
TITLE	Yu,J.X., Chao,L. and Chao,J.			
JOURNAL	Molecular cloning, tissue-specific expression, and cellular			
MEDLINE	localization of human prostaticin mRNA			
PUBMED	J Biol. Chem. 270 (22), 13483-13489 (1995)			
REFERENCE	95286644			
AUTHORS	2 (bases 1 to 7008)			
TITLE	7768952			
JOURNAL	Yu,J.X., Chao,L., Ward,D.C. and Chao,J.			
MEDLINE	Structure and chromosomal localization of the human prostaticin			
PUBMED	(PRSS8) gene			
REFERENCE	Genomics 32 (3), 334-340 (1996)			
AUTHORS	96435910			
TITLE	8838796			
JOURNAL	3 (bases 1 to 7008)			
MEDLINE	Yu,J.X., Chao,L., Ward,D.C. and Chao,J.			
PUBMED	Direct Submission			
REFERENCE	Submitted (07-AUG-1995) Jack X Yu, Biochemistry and Molecular			
AUTHORS	Biology, Medical University of South Carolina, 171 Ashley Avenue,			
TITLE	Charleston, SC 29425, USA			
JOURNAL	Location/Qualifiers			
MEDLINE	1..7008			
PUBMED	FEATURES			
REFERENCE	source			

**mRNA**

```

/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="16"
/map="16p11.2"
join(1525, .1839, 2082, .2099, 3863, .4025, 4297, .4563
.6654, .4820, 4913, .5811)

```

BASE COUNT	1435	a	1991	c	2064	g	1518	t
exon	/product="prostasin"							
	1525. .1838							
	/number=1							
cds	join(1754. .1838,2082. .2099,1863. .4025,4297. .4568,							
	4654. .4820,4913. .5239)							
	/function="serine proteinase"							
	/codon_start=1							
	/product="prostasin"							
	/protein_id="AAB19071.1"							
	/db_xref="GI:1143194"							
	/translation="MAQKGVLPQGQAVAILLYLGLIRGCTGAEGABAPCGVAPQAR							
	ITDSSSAVQAQMPQVSIITIEGVAVCGSLVSEQWVLSAAHCFSEHHKEAYEVKLG							
	HOLDSSVEDAKVSTLKDIIIPHSYLOGSGDIALDLSRPITPSRYIRPICLPAAANA							
	SPNGHGHCTVTWGVAHPASVSLITPKRLOOLEVLIISRETCNCIYNTDAKREEHFVC							
	EDWACAGYVGGKDACQDGSQSGPASCVEBGMITGTIVSNGDACANNRPQVYTLASS							
	YAWMISKVTLEQPRVVPQTOESQPSDNLGSHLAFSSAPAQGLRPILFLPIALAG							
	LISPMTLSEH"							
exon	2082. .2099							
	/number=2							
exon	3863. .4025							
	/number=3							
exon	4297. .4568							
	/number=4							
exon	4654. .4820							
	/number=5							
exon	4913. .5811							
	/number=6							

Query Match	52.7%	Score 879.2	DB 9	Length 7008
Best Local Similarity	84.2%	Pred. No. 9,9e-167		
Matches 1158; Conservative	3	Mismatches 29	Indels 185	Gaps 8

QY	472	CAGGAGCACCACAAAGGAACCTTAAGAGGTCAAGCTGGGGGGCCACAGACTAGACTCTCTA	531
Db	4294	CAGGAGCACCACAAAGGAACCTTAAGAGGTCAAGCTGGGGGGCCACAGACTAGACTCTCTA	4353
QY	532	CTCCGAGAGAGCCAAAGTTCAGCAACCTTGAAGAGACATCATCCCCACCCAGCTACTTCCA	591
Db	4354	CTCCGAGAGAGCCAAAGTTCAGCAACCTTGAAGAGACATCATCCCCACCCAGCTACTTCCA	4413
QY	592	GGAAGGCTCCCAAGGGCGACATTGCACTCTTCCAACTCAGAGAGCCATCATCTTCTCCCG	651
Db	4414	GGAAGGCTCCCAAGGGCGACATTGCACTCTTCCAACTCAGAGAGCCATCATCTTCTCCCG	4473
QY	652	CTTATATCCCGGCCCATCTGCTCTCCCTGAGAGCAAGCCTCTTCTCCCAAGGCTTCCACTG	711
Db	4474	CTTATATCCCGGCCCATCTGCTCTCCCTGAGAGCAAGCCTCTTCTCCCAAGGCTTCCACTG	4533
QY	712	CACGTGTACATGAGCTGGGGGTCAATGGGCCCCCT-----	743
Db	4534	CACGTGTACATGAGCTGGGGGTCAATGGGCCCCCT-----	4593
QY	744	-----CAG 746	
Db	4594	GGTGTGAGAGGGGCACTGACTTGGGGAGGGGCCAGGGTAAAGCTCTTTAAACCCCAAG	4653
QY	747	TGAGCTCTCTGACAGCCCAAGCCACTGACGACACTCGAGGTGCTCTGTATCATGTGTGAGA	806
Db	4654	TGAGCTCTCTGACAGCCCAAGCCACTGACGACACTCGAGGTGCTCTGTATCATGTGTGAGA	4713
QY	807	CGTGTATACCTGCTGTATCAACATTCGAGAGCCCAAGGCTTGAAGAGCGCACTTTGTCCAGAG	866
Db	4714	CGTGTATACCTGCTGTATCAACATTCGAGAGCCCAAGGCTTGAAGAGCGCACTTTGTCCAGAG	4772
QY	867	GACATGATGTGTGCTGTGCTATGTGAGAGGGGGGCAAGAGCGCTTGC-----	911
Db	4773	GACATGATGTGTGCTGTGCTATGTGAGAGGGGGGCAAGAGCGCTTGC-----	4832
QY	912	-----	911
Db	4833	CCGGGGCAGATGACCAAGTGCACCTTCGAAAGAGAGGCTGAGCCGGTCTGTATGCTGC	4892

Qy	912	-----	CAGGGTGA	CTCTGAGGAGCCCACTCTCCGACCTTGGAGGGTCC	954
Db	4893	TGTGTGGCTTCTCTCTC	CAGGGTGA	CTCTGAGGAGCCCACTCTCCGACCTTGGAGGGTCC	4952
Qy	955	TCTGTGTA	CTGACGGGCA	TTGTGAGCTGGGGAGATGCTGTGTGGGACCCGCAACAGGCTGTG	1014
Db	4953	TCTGTGTA	CTGACGGGCA	TTGTGAGCTGGGGAGATGCTGTGTGGGACCCGCAACAGGCTGTG	5012
Qy	1015	GTGTGTAC	CTCTGGCCCTCCAG	CTATGCTCTGTGATTCGAAGCAAGGTGACAAATCTCC	1074
Db	5013	GTGTGTAC	CTCTGGCCCTCCAG	CTATGCTCTGTGATTCGAAGCAAGGTGACAAATCTCC	5072
Qy	1075	AGCCTCGTGTGATG	AGCCCAAA	CCAGAGAGTCCAGCCCGACAGCAAACTCTGTGTGAGAGCC	1134
Db	5073	AGCCTCGTGTGATG	AGCCCAAA	CCAGAGAGTCCAGCCCGACAGCAAACTCTGTGTGAGAGCC	5132
Qy	1135	ACCTGTGCTT	CAGCTCTG	CCCCAGCCCGAGGCTTGTGAGGACCACTCTTTTCTGCTCTC	1194
Db	5133	ACCTGTGCTT	CAGCTCTG	CCCCAGCCCGAGGCTTGTGAGGACCACTCTTTTCTGCTCTC	5192
Qy	1195	TGGGCTTGTGCTT	CTGGGCTCTCTCTT	CCCATATGGCTCAGGAGACATCTGAGCTTGGCTTACTT	1254
Db	5193	TGGGCTTGTGCTT	CTGGGCTCTCTCTT	CCCATATGGCTCAGGAGACATCTGAGCTTGGCTTACTT	5252
Qy	1255	CCAGAGATG	ATGACATCA	CACTCAAGAGACAGAGCCTGTGCTTCCCTGATGGCCTTTTGA	1314
Db	5253	CCAGAGATG	ATGACATCA	CACTCAAGAGACAGAGCCTGTGCTTCCCTGATGGCCTTTTGA	5312
Qy	1315	CCCAAGGCTT	GACCTT	GTGAGCCACTCTCCCTTCAAGACTCTGCGGAGGCTGGGACCCCAT	1374
Db	5313	CCCAAGGCTT	GACCTT	GTGAGCCACTCTCCCTTCAAGACTCTGCGGAGGCTGGGACCCCAT	5372
Qy	1375	CTTATATCTT	TAGCCCATCTTCTT	GTGGGAGTCTTTTGGGACATCACTGAGATGACGA	1434
Db	5373	CTTATATCTT	TAGCCCATCTTCTT	GTGGGAGTCTTTTGGGACATCACTGAGATGACGA	5432
Qy	1435	GTTTTACTG	CTGTAGCAATG	AGGACGCTCTGGCCCTTCAACACCATGAGACGACCC	1494
Db	5433	GTTTTACTG	CTGTAGCAATG	AGGACGCTCTGGCCCTTCAACACCATGAGACGACCC	5492
Qy	1495	ATTGGGCGAGNT	CCTGGGGAG	-TCTTGGGACCTTGTATGAAATGAGCCCTGTGGGTTT	1552
Db	5493	ATTGGGCGAGNT	CCTGGGGAG	-TCTTGGGACCTTGTATGAAATGAGCCCTGTGGGTTT	5551
Qy	1553	CCACCTGTTT	CTMGAAAG	CTGTCCCGGCCGCTTCCGAGCTMGATGAGCACTTT	1612
Db	5552	CCACCTGTTT	CTMGAAAG	CTGTCCCGGCCGCTTCCGAGCTMGATGAGCACTTT	5607
Qy	1613	TTTTGGCCNTT	CCCCGCTGTTT	TGGGCTTGGGCACTTTTGAAGTTTGAAGAGA	1667
Db	5608	CTGTGCCCT	CTTCCCTGTGTTT	CTGGGCTGGGACCTTTGTGACGCTTGAGAGACA	5662



```

/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="16"
/clone="CMD-251B20"
/clone_lib="Caltech human BAC library D"
BASE COUNT 44605 a 52177 c 51642 g 44918 t 2140 others
ORIGIN

```

Query Match	52.6%	Score 877.6	DB 2	Length 195476
Best Local Similarity	84.3%	Pred. No. 1.1e-166		
Matches 1156, Conservative	3	Mismatches 25	Indels 187	Gaps 8

QY	472	CAGGAGACACCAAGAAGCCTATGAGGCAAGCTGAGGGGCCCAACAGCTAACCTCTTA	531
Db	82226	CAGGAGACACCAAGAAGCCTATGAGGCAAGCTGAGGGGCCCAACAGCTAACCTCTTA	82267
QY	532	CTCCGAGAGCGCAAGGTCAAGACCCCTGAAGGAGCATCATCCCCACCCAGTCACTCTCA	591
Db	82266	CTCCGAGAGCGCAAGGTCAAGACCCCTGAAGGAGCATCATCCCCACCCAGTCACTCTCA	82267
QY	592	GGAGGGCTCCGAGGGCGCAATTGCACTCTTCCAACTCAGAGAACCATCACTTCTCCG	651
Db	82206	GGAGGGCTCCGAGGGCGCAATTGCACTCTTCCAACTCAGAGAACCATCACTTCTCCG	82147
QY	652	CTACATCCGGCCCATCTGCGCTCCCTGAGCAAGCGCTCTTCCCAAGCGCTCCCACTG	711
Db	82146	CTACATCCGGCCCATCTGCGCTCCCTGAGCAAGCGCTCTTCCCAAGCGCTCCCACTG	82087
QY	712	CACCTGTCACTGAGCTGGGGTCAATGGCCCCCT	743
Db	82086	CACCTGTCACTGAGCTGGGGTCAATGGCCCCCTCAGGTGAGTGGAGCGCTGGTGCCTTA	82027
QY	744	-----CA	745
Db	82026	AGGTGTGAGGGGGCACCTGTGACTTGGGGAGGGGCCCAAGGTAAAGCTCTTTTACCCCA	81967
QY	746	GTGAGGCTCCTGACGGCCCAAGGCCACTGACAGCACTGAGTGGCTCTGATCACTGCTGAG	805
Db	81966	GTGAGGCTCCTGACGGCCCAAGGCCACTGACAGCACTGAGTGGCTCTGATCACTGCTGAG	81907
QY	806	ACGTGTGAACTGCTGCTGATCAACAATCAGACGCAAGCTGAGAGACCCCACTTGTCCAA	865
Db	81906	ACGTGTGAACTGCTGCTGATCAACAATCAGACGCAAGCTGAGAGACCCCACTTGTCCAA	81848
QY	866	GGACATGATGTGTGTGCTGCTATGTGAGGGGGGCAAGACGCTTGCA	913
Db	81847	GGACATGATGTGTGTGCTGCTATGTGAGGGGGGCAAGACGCTTGCAAGACAG	81788
QY	914	-----	913
Db	81787	CCCGGAGGCGAGATGACCAAGTGCMACTTGGAAAGAGAGCGCTGAGCT	81728
QY	914	-----GGGTGACCTCTGGGGGCCCACTCTCTGCGCTGAGGAG	951
Db	81727	GCTGTGTGGCTTCTCTCTCAGGGTTGACTCTTGGGGGCCCACTCTCTGCGCTGAGGAG	81668
QY	952	GTCTCTGTGATCCTGACGGGCAATTGTGAGCTGGGAGATGCTTGTGGGGCCCGACAGC	1011
Db	81667	GTCTCTGTGATCCTGACGGGCAATTGTGAGCTGGGAGATGCTTGTGGGGCCCGACAGC	81608
QY	1012	CTGGTGTGTATCACTCTGCGCTCCAGCTATAGCTCTCTGATCCAAAGCAAGTGAAGATC	1071
Db	81607	CTGGTGTGTATCACTCTGCGCTCCAGCTATAGCTCTCTGATCCAAAGCAAGTGAAGATC	81548
QY	1072	TCCAGGCTCGTGTGTGTGCCCAAGCCAGAGATCCAGGCGCGACAGCACTCTGTGGCA	1131
Db	81547	TCCAGGCTCGTGTGTGTGCCCAAGCCAGAGATCCAGGCGCGACAGCACTCTGTGGCA	81488
QY	1132	GCCACTTGAGCTTCAAGCTTGCCCAAGCCAGGGCTTGTGTAGAGCCCACTTTTCTTC	1191
Db	81487	GCCACTTGAGCTTCAAGCTTGCCCAAGCCAGGGCTTGTGTAGAGCCCACTTTTCTTC	81428
QY	1192	CTCTGGGCGTGGCTCTGGGCTCTCTCCCAATGCTCAGCGAGCACTAGCTGGCGCTTA	1251

Db	81427	CTCTGGGCGCTGGCTCTGGGCGCTCTCTCTCCCGCATGGCTGACGAGCACTAGCTGGCGCTTA	81366
Qy	1252	CTTCCAGATGATGATGCATCACACTCAAGACAGAGAGCGCTGTCCTTCCCTGATGGCTTT	1311
Db	81367	CTTCCAGATGATGATGCATCACACTCAAGACAGAGAGCGCTGTCCTTCCCTGATGGCTTT	81308
Qy	1312	GGACCCAGAGGCGCTGACTTGAAGCCACTCTCTTCTTCAAGACTCTGGGGAGGCTGGGGCCC	1371
Db	81307	GGACCCAGAGGCGCGTGAAGCCACTCTCTTCTTCAAGACTCTGGGGAGGCTGGGGCCC	81248
Qy	1372	CATCTTGATCTTTGAGCCCATCTCTTGAGTGTGCTTTTGGAGCCATACAGAGTCA	1431
Db	81247	CATCTTGATCTTTGAGCCCATCTCTTGAGTGTGCTTTTGGAGCCATACAGAGTCA	81188
Qy	1432	GGAGTTTACTGCTGTAGCAATGAGCCAGAGCTCTGGCGCCCTGACACCATAGAGCCAG	1491
Db	81187	GGAGTTTACTGCTGTAGCAATGAGCCAGAGCTCTGGCGCCCTGACACCATAGAGCCAG	81128
Qy	1492	CCCATTTGSSCGAGNCTCTGGGAG--TCTGGAGACTTTGGTATGAAATGAGCCCTGGG	1549
Db	81127	CCCATTTGSSCGAGCTCTGGGAGAGCTCTGGGAGCCCTTGGTATGAAATGAGCCCT--GG	81069
Qy	1550	TTCGCCACTGTTCTTNGAAGACTGCTTCCCGGCGCGGCTTCCCAAGCTNGATGAGACACT	1603
Db	81068	CTTCCCACTGTTCTTNGAAGACTGCT--TCCCGGCGCGGCTGCCAAGCT--GATGAGACACA-	81012
Qy	1610	TTTTTTTGGCMTTCCCTGCTGTTTTTGGGTTGGGCACTTTTGGAGTTT	1660
Db	81011	TCTCTCTGCGCTCTCCCTCTGTGTTGTGGAGCTGGGGGCACTTTTGTGCAAGCTT	80961

[illegible]

ACCESSION	AX675583
VERSION	AX675583.1
KEYWORDS	GI:29333568

SOURCE	Homo sapiens (human)
ORGANISM	Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

## REFERENCE AUTHORS

**AUTHORS**  
Padigkar, M., Li, L., Zernussen, B.D., Casman, S.J., Shenoy, S.,  
Spytek, K.A., Zhong, M., Gangoli, E.A., Burgess, C.E., Patirajan, M.,  
Votaw, C.A., Taylor, S., Tchernov, V.T., Miller, C.E., Guo, X.,  
Balogh, F.L., Grosse, W.M., Alsbrook, J.P., Gerlach, V.,  
Edingermark, S., Rothenberg, M.E., Ellerman, K., MacDougall, J.,  
Maljanek, U., Millet, I., Feyman, O., Smithson, G., Gunther, E., and  
Stone, D.J.

TITLE	JOURNAL
Proteins, polynucleotides encoding them and methods of using the same	WO 02055704-A 33 18-JUL-2002; Curagen Corporation (US)

## FEATURES

Location/Qualifiers  
1. .882

```

/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
BASE COUNT      165 a      304 c      254 g      159 t
ORIGIN

```

Query Match	Score	DB	Length
51.4%	857.4	6	882

Best Local Similarity 99.5%; Freq. NO. 3.0e-16z;  
Matches 869; Conservative 1; Mismatches 2; Indels 1; Gaps 1;

QY 296 GGGAGAGGGGCGAGAGCTTCCTGGGTGTGGCCCCCAGCAGCCATCACAAGGTGGAGC 355  
 :  
 Db 7 GGGAGAGGGGCGAGAGCTTCCTGGGTGTGGCCCCCAGCAGCCATCACAAGGTGGAGC 66  
 QY 356 AGTGAAGTCCGCGTCATGTGGCCCTGGCAGGTACAGTCACTATGAAGCGTTCATGTG 415

Db	67	AGTGCAGTCCGCGGTCAGTGGCCCTGGCAGGTCAGATCACTATGAAGCGTCCAGTGTG	126
QY	416	TGTGTGGCTCTCTGTGTCTGAGCAGTGGGTGTGTGACGTCTCACTGTCTTCCCAAGC	475
Db	127	TGTGTGGCTCTCTGTGTCTGAGCAGTGGGTGTGTGACGTGTGTCACTCTTCCCAAGC	186
QY	476	GAGCACCAGAAAGAGCTATGAGTGCATGAGTGGGGGCCCAACGAGCTAGACTCTACTTCC	535
Db	187	GAGCACCAGAAAGAGCTATGAGTGCATGAGTGGGGGCCCAACGAGCTAGACTCTACTTCC	246
QY	536	GAGCACCAGAGTCAAGCACCCTGAGAGACATCATCCCAACCCCAAGCTATCTTCAGAG	595
Db	247	GAGCACCAGAGTCAAGCACCCTGAGAGACATCATCCCAACCCCAAGCTATCTTCAGAG	306
QY	596	GGCTTCCAGGGCGACATTGCACTCTTCAACTCAGCAGACCCATCACTTCTCCCGTAC	655
Db	307	GGCTTCCAGGGCGACATTGCACTCTTCAACTCAGCAGACCCATCACTTCTCCCGTAC	366
QY	656	ATCCGGCCCATCTGCTCCCTCCGACGACCAAGCCCTCTCCCAAGGGCTCCACTGACT	715
Db	367	ATCCGGCCCATCTGCTCCCTCCGACGACCAAGCCCTCTCCCAAGGGCTCCACTGACT	426
QY	716	GTCACTGGCTGGGGTCAATGTGAGCCCTCACTGAGCCTCTGAGCGCCCAAGCCACTGAG	775
Db	427	GTCACTGGCTGGGGTCAATGTGAGCCCTCACTGAGCCTCTGAGCGCCCAAGCCACTGAG	486
QY	776	CAACTCGAGTGCCTCTGATATCATGTGTGAGACGTGTGTACTGCTGTTCACATCGACGC	835
Db	487	CAACTCGAGTGCCTCTGATATCATGTGTGAGACGT- GTAACCTGCTGTATCAACATCGACGC	545
QY	836	CAAGCTGAGAGGCGGCACTTTGTTCAAAGAGACATGGTGTGTGCTGAGTGTGAGAGGG	895
Db	546	CAAGCTGAGAGGCGGCACTTTGTTCAAAGAGACATGGTGTGTGCTGAGTGTGAGAGGG	605
QY	896	GAGCAGAGACCGCTTCCCAAGGTTGACTCTGGGGGCCCACTCTCTGCTGTGAGGGTCT	955
Db	606	GAGCAGAGACCGCTTCCCAAGGTTGACTCTGGGGGCCCACTCTCTGCTGTGAGGGTCT	665
QY	956	CTGTGACTCTGACGGGCAATTGTGAGCTGGGGAGATGCTGTGGGGCCGCAACAGGCTGG	1015
Db	666	CTGTGACTCTGACGGGCAATTGTGAGCTGGGGAGATGCTGTGGGGCCGCAACAGGCTGG	725
QY	1016	TGTGTACACTCTGGGCTCAGCTACGTAAGCTCTGAGTCCAAAGCAAGGTGACAGAACTCCA	1075
Db	726	TGTGTACACTCTGGGCTCAGCTACGTAAGCTCTGAGTCCAAAGCAAGGTGACAGAACTCCA	785
QY	1076	GCTTGTGTGTGGTCCCAAAACCCAGAGATGCCAGGCCGACAGCAACCTGTGTGACAGCA	1135
Db	786	GCTTGTGTGTGGTCCCAAAACCCAGAGATGCCAGGCCGACAGCAACCTGTGTGACAGCA	845
QY	1136	CCTGGCTTTCAGACTCTGCTGCCCAAGCCCAAGGGCTT	1168
Db	846	CCTGGCTTTCAGACTCTGCTGCCCAAGCCCAAGGGCTT	878
RESULT 15			
LOCUS	AR219285	1142 bp	DNA
DEFINITION	Sequence 8 from patent US 6420157.	linear	PAT 25-SEP-20020
ACCESSION	AR219285		
VERSION	AR219285.1	GI:23320255	
KEYWORDS			
SOURCE	Unknown.		
ORGANISM	Unknown.		
REFERENCE	1 (bases 1 to 1142)		
AUTHORS	Darrow,A., Qi,J. and Andrade-Grodon,P.		
TITLE	Zymogen activation system		
JOURNAL	Patent: US 6420157-A 8 16-JUL-2002;		
FEATURES	Location/Qualifiers		
source	1..1142	/organism="unknown"	
BASE COUNT	235 a	360 c	303 g 244 t

[illegible]

Search completed: December 15, 2003, 22:29:52  
Job time : 6331 secs

REFERENCE	1 (bases 1 to 1142)	
AUTHORS	Darrow, A., Qi, J. and Andrade-Groden, P.	
TITLE	Zymogen activation system	
JOURNAL	US 6420157-A 8 16-JUL-2002;	
FEATURES	Location/Qualifiers	
source	1..1142	
	/organism="unknown"	
BASE COUNT	235 a 360 c 303 g 244 t	



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 15, 2003, 19:58:57 / Search time 3746 Seconds  
(without alignments)  
10822.175 Million cell updates/sec

Title: US-09-925-301-208

Perfect score: 1668  
Sequence: 1 cacactgcctcgcctcgata.....tttggaaagttcaggagaa 1668

Scoring table: IDENTITY\_NUC  
Gapop 10.0, Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

EST:\*  
1: em\_estba:\*  
2: em\_estchum:\*  
3: em\_estin:\*  
4: em\_estmu:\*  
5: em\_estrov:\*  
6: em\_estropl:\*  
7: em\_estro:\*  
8: em\_hlc:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_hrc:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: em\_gss\_hum:\*  
18: em\_gss\_hiv:\*  
19: em\_gss\_pln:\*  
20: em\_gss\_vrt:\*  
21: em\_gss\_fun:\*  
22: em\_gss\_mam:\*  
23: em\_gss\_mus:\*  
24: em\_gss\_pro:\*  
25: em\_gss\_rtd:\*  
26: em\_gss\_phg:\*  
27: em\_gss\_vrl:\*  
28: gb\_gss1:\*  
29: gb\_gss2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	908.4	54.5	1071	9	AL551470
2	849.8	50.9	888	14	CA489381
3	839.8	50.3	1013	12	BM923713
4	816.8	49.0	865	14	CD389322

Result No.	Score	Query Match	Length	DB ID	Description
5	810.8	48.6	922	13	BQ690230
6	795.2	47.7	909	14	CD050808
7	769.4	46.1	861	12	BT763558
8	762.2	45.7	815	14	CB995955
9	753.2	45.2	1057	13	CB386100
10	751.4	45.0	1088	9	AL575026
11	743.6	44.6	1201	13	BX386211
12	743.4	44.6	768	13	BQ482196
13	741	44.4	765	10	BQ482196
14	738.4	44.3	748	12	BT766651
15	732.4	43.9	840	14	CA487658
16	730.4	43.8	810	14	CB996835
17	723.6	43.4	790	9	AU123750
18	720.2	43.2	786	10	BQ481339
19	719	43.1	960	13	BQ691278
20	713.4	42.8	842	14	CB992867
21	711	42.6	1629	11	AK010640
22	710.4	42.6	831	14	CB996348
23	705.8	42.3	1699	11	AK078696
24	704.6	42.2	897	13	BQ689953
25	695	41.7	772	14	CB961615
26	688.2	41.3	741	9	AI978814
27	687.6	41.2	768	14	CB995671
28	684.8	41.1	694	13	BQ691586
29	677.6	40.6	699	14	CB851297
30	676.6	40.6	772	9	AU141216
31	668.8	40.1	1201	13	BX387681
32	663.6	39.8	690	10	BQ481221
33	661	39.6	804	10	BQ435445
34	654.6	39.2	683	2	HSW066452
35	648.6	38.9	745	10	BG743637
36	637.8	38.2	657	9	AI935459
37	636	38.1	796	10	BG386313
38	635.6	38.1	715	10	BE989650
39	633	37.9	655	12	BM781586
40	632.4	37.9	1201	13	BX386210
41	630.4	37.8	901	13	BU185512
42	627	37.6	705	9	AU135075
43	623.8	37.4	742	9	AU142128
44	623.2	37.4	670	9	AU121662
45	618.2	37.1	619	12	BM794592

## ALIGNMENTS

RESULT 1  
AL551470 1071 bp mRNA EST 31-MAY-2003  
LOCUS AL551470 Homo sapiens PLACENTA COR 25-NORMALIZED Homo sapiens cDNA  
DEFINITION clone CS0D1064YF02 5-PRIME, mRNA sequence.  
ACCESSION AL551470.2 GI:31273286  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM Homo sapiens (human)  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE  
AUTHORS Li, W.B., Gruber, C., Jessee, J., and Polajars, D.  
TITLE Full-length cDNA libraries and normalization  
JOURNAL Unpublished  
COMMENT On Feb 15, 2001 this sequence version replaced gi:12889449.

Contact: Genoscope - Centre National de Sequencage  
Genoscope - Centre National de Sequencage  
BP 191 91006 Evry cedex - France  
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr  
Library was constructed by life technologies, a division of  
Invitrogen. This sequence belongs to sequence cluster 7896.r for  
more information about this cluster, see  
http://www.genoscope.cns.fr/  
cgi-bin/cluster.cgi?seq=CS0D1064DC01QPlac1uster=7896.r. Contact :  
Feng liang Email : fliang@lifetech.com URL :

http://fulllength.invitrogen.com/ Invitrogen Corporation 1600  
 Faraday Avenue Genoscope Sequence ID : CS0D1064DC01QPL  
 Location/Qualifiers

FEATURES  
 source

1..1071  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="CS0D10641P02"  
 /issue\_type="PLACENTA COT 25-NORMALIZED"  
 /clone\_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"  
 /note="1st strand cDNA was primed with a NotI-oligo(dT)  
 primer. Five prime and enriched, double-strand cDNA was  
 digested with Not I and cloned into the Not I and EcoR V  
 sites of the pCMVSPORT 6 vector. Library was normalized."

BASE COUNT 200 a 345 c 315 g 206 t 5 others

Query Match 54.5%; Score 908.4; DB 9; Length 1071;  
 Best Local Similarity 97.1%; Pred. No. 2,1e-181;  
 Matches 963; Conservative 4; Mismatches 20; Indels 5; Gaps 4;

124 ACATCTGGGCTCTGCTTGCACACCAACCCAGGCTTCTACCTTGGCTGAGTCT 183  
 63 ATATCTGGGCTCTGCTTGCACACCAACCCAGGCTTCTACCTTGGCTGAGTCT 122  
 184 GCCCAGGGGCTTGTCTTGGCCATGCGCAGAGGGGCTCTGGGGCTGGGAGCT 243  
 123 GCCCAGGGGCTTGTCTTGGCCATGCGCAGAGGGGCTCTGGGGCTGGGAGCT 182  
 244 GGGGGCTGTGGCCATCTGCTATCTTGTGATTACTCCGGTCGGGGACAGAGCGGAGG 303  
 183 GGGGGCTGTGGCCATCTGCTATCTTGTGATTACTCCGGTCGGGGACAGAGCGGAGG 242  
 304 GGCAGAGCTCTCTGCGGTGGCGCCCAAGACGATCAAGGTGGACAGTGCAGT 363  
 243 GGCAGAGCTCTCTGCGGTGGCGCCCAAGACGATCAAGGTGGACAGTGCAGT 302  
 364 GCGCGCTCAAGTGGCGCCCTGGAGGATCACTATGAAGGGCTCCATGTGTGTGG 423  
 303 GCGCGCTCAAGTGGCGCCCTGGAGGATCACTATGAAGGGCTCCATGTGTGTGG 362  
 424 CTCTCTGTCTGAGGAGTGGGTGTGTGAGTGTCTCACTGTCTCCCGACAGACCA 483  
 363 CTCTCTGTCTGAGGAGTGGGTGTGTGAGTGTCTCACTGTCTCCCGACAGACCA 422  
 484 CAAAGAGCTTATAGGTCAAGCTGGGGGCGCCACAGCTAGACTCTTCTCCAGAGCGC 543  
 423 CAAAGAGCTTATAGGTCAAGCTGGGGGCGCCACAGCTAGACTCTTCTCCAGAGCGC 482  
 544 CAAAGTCAAGCCCTGAAAGACATCATCCCCACCCAGCTACTCCAGAGGGGCTCCA 603  
 483 CAAAGTCAAGCCCTGAAAGACATCATCCCCACCCAGCTACTCCAGAGGGGCTCCA 542  
 604 GGGGCACTTATGCACTCTCCAACTCAGAGAGCCATCACTTCTCCGCTACATCCGCGC 663  
 543 GGGGCACTTATGCACTCTCCAACTCAGAGAGCCATCACTTCTCCGCTACATCCGCGC 602  
 664 CATCTGCTCTCCCTCAGGCAAGCGCTCTTCCCAAGGGCTCTCACTGCACTGTCACTGG 723  
 603 CATCTGCTCTCCCTCAGGCAAGCGCTCTTCCCAAGGGCTCTCACTGCACTGTCACTGG 662  
 724 CTGGGGTATATGGCCCCCTTCAAGAGCTCTCTGACGCCCAAGGCACTGACAGCACTTGA 783  
 663 CTGGGGTATATGGCCCCCTTCAAGAGCTCTCTGACGCCCAAGGCACTGACAGCACTTGA 721  
 784 GGTGCTCTGATCTAGTGTGAGAGGTGTAATCTGCTTACAAACATTCAGACCCCAAGCTTG 843  
 722 GGTGCTCTGATCTAGTGTGAGAGGTGTAATCTGCTTACAAACATTCAGACCCCAAGCTTG 780  
 844 AGAGAGCCGCACTTTGTCCAAAGAGCATGTGTGTGTGCTGCTATGTGAGAGGGGGCAAGG 903  
 781 AGAGAGCCGCACTTTGTCCAAAGAGCATGTGTGTGTGCTGCTATGTGAGAGGGGGCAAGG 840

QY 904 AGCGCTGCCAGGGTGAATCTTGGGGGCGCCACTCTCTCTGCTGTGAGGGTCTCTGTACC 963  
 DB 841 AGCGCTGCCAGGGTGAATCTTGGGGGCGCCACTCTCTCTGCTGTGAGGGTCTCTGTACC 900  
 QY 964 TGACGGGCACTTGTGAGCTGGGGAGATGCTGTGGGGCGCCGCAAGCGCTGGTGTCA 1023  
 DB 901 TGACGGGCACTTGTGAGCTGGGGAGATGCTGTGGGGCGCCGCAAGCGCTGGTGTCA 960  
 QY 1024 CTCTGGGCTCCGAGTATGCTCTCTGATCAAGAGCAAGGTGACAGACTCCAGCTCTGTG 1083  
 DB 961 CTCTGGGCTCCGAGTATGCTCTCTGATCAAGAGCAAGGTGACAGACTCCAGCTCTGTG 1017  
 QY 1084 TGGTCCCCCAACCCAGAGATGCCAGCCCGAC 1115  
 DB 1018 TGGTCCCCCAACCCAGAGATGCCAGCCCGAC 1049

RESULT 2  
 CA489381 888 bp mRNA linear EST 14-NOV-2002  
 LOCUS  
 DEFINITION  
 ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Homo sapiens (human)  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 COMMENT  
 Email: cgabbs-r@mail.nih.gov  
 Tissue Procurement: Kristi A. Eglund, Ira Pastan  
 CDNA Library Preparation: Invitrogen Corp  
 CDNA Sequencing by: The I.M.A.G.E. Consortium (LNL)  
 DNA Sequencing by: Agencourt Bioscience Corporation  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LNL at:  
 http://image.llnl.gov  
 Plate: L1AM14283 row: h column: 17  
 High quality sequence stop: 701.

# FEATURES

source

1..888  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:6721793"  
 /cell\_line="ZR-75-1, MCF7, SK-BR-3, MDA-MB-231, HTERT-HME1  
 LNCaP"  
 /lab\_host="EMDH10B"  
 /clone\_lib="MABCL"  
 /note="Vector: pCMV-SPORT6, Site 1: EcoRV, Site 2: Not I;  
 Subtracted with brain, liver, lung, kidney and muscle.  
 directionally cloned. Priming method: oligo-dT. Average  
 insert size: 1800 bp. Library amplification: 26,000 fold.  
 Kristi A. Eglund, James J. Vincent, Robert Strauberg,  
 Bungkook Lee & Ira Pastan: Discovery of new breast  
 cancer genes encoding membrane and secreted proteins.  
 Manuscript submitted."

BASE COUNT 161 a 321 c 238 g 168 t

Query Match 50.9%; Score 849.8; DB 14; Length 888;  
 Best Local Similarity 99.4%; Pred. No. 4.9e-169;  
 Matches 884; Conservative 0; Mismatches 2; Indels 3; Gaps 3;

QY 467 TTCCCCAGGAGCAACCAAGGAAGCTTATGAGTCAAGCTGGGGGCGCCACCGCTGAC 526  
 DB 1 TTCCCCAGGAGCAACCAAGGAAGCTTATGAGTCAAGCTGGGGGCGCCACCGCTGAC 60

QY 527 TCTACTCCGAGACGCCAAGGTACGACCTTGAGGACATCATCCCAACCCCAAGTAC 586  
 Db 61 TCTACTCCGAGACGCCAAGGTACGACCTTGAGGACATCATCCCAACCCCAAGTAC 120  
 QY 587 CTCGAGAGGGGCTCCCAAGGGGACATTCGACTCTCTCCAACTCCAGACCAACCATCACTTC 646  
 Db 121 CTCGAGAGGGGCTCCCAAGGGGACATTCGACTCTCTCCAACTCCAGACCAACCATCACTTC 180  
 QY 647 TCCCGCTACATCCGAGCCCATCTGCTCCCTCCGACGCAACGCTCTCTCCCAAGGCTC 706  
 Db 181 TCCCGCTACATCCGAGCCCATCTGCTCCCTCCGACGCAACGCTCTCTCCCAAGGCTC 240  
 QY 707 CACTGCACTGTCACTGCTGCGGCTCATGTGCGCCCTCACTGAGCCCTCTGACGCCAAG 766  
 Db 241 CACTGCACTGTCACTGCTGCGGCTCATGTGCGCCCTCACTGAGCCCTCTGACGCCAAG 300  
 QY 767 CCACTGAGAGAACTCCGAGGCTCTGCTCTGATCATGTCTGAGACGTGTAACTGCTGCTCA 826  
 Db 301 CCACTGAGAGAACTCCGAGGCTCTGCTCTGATCATGTCTGAGACGTGTAACTGCTGCTCA 359  
 QY 827 CATGACGCGCAAGCTGAGAGCGCACTTGTCTCAAGAGACATGCTGTGCTGCTGCTCA 886  
 Db 360 CATGACGCGCAAGCTGAGAGCGCACTTGTCTCAAGAGACATGCTGTGCTGCTGCTCA 419  
 QY 887 TGTGAGAGGGGGGCAAGAGCGCTGCGAGGTGACTCTGCGGCGCCCACTCTCTGCTGCT 946  
 Db 420 TGTGAGAGGGGGGCAAGAGCGCTGCGAGGTGACTCTGCGGCGCCCACTCTCTGCTGCT 479  
 QY 947 GAGAGGCTCTTGTGATCCTGACCGGGCATTTGAGAGTGGGGAGATGCTCTGCGGCGCC 1006  
 Db 480 GAGAGGCTCTTGTGATCCTGACCGGGCATTTGAGAGTGGGGAGATGCTCTGCGGCGCC 539  
 QY 1007 CAGGCTGTGTGTGATCCTGAGCGCTCAGCTATGCTCTCCGATCCAAAGCAAGGTAC 1066  
 Db 540 CAGGCTGTGTGTGATCCTGAGCGCTCAGCTATGCTCTCCGATCCAAAGCAAGGTAC 599  
 QY 1067 AGAATCTCAGACCTGTGTGTGTCGCCCAACCCAGAGATCCCAAGCGGACCAACTCTG 1126  
 Db 600 AGAATCTCAGACCTGTGTGTGTCGCCCAACCCAGAGATCCCAAGCGGACCAACTCTG 659  
 QY 1127 TGGCAGCAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1185  
 Db 660 TGGCAGCAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 719  
 QY 1186 TCTGCTCTGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1245  
 Db 720 TCTGCTCTGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 779  
 QY 1246 GGCCTACTTCAGAGATGATGATCACTCAAGAGACAGAG-CTGTGCTCTTCCCTGAT 1304  
 Db 780 GGCCTACTTCAGAGATGATGATCACTCAAGAGACAGAG-CTGTGCTCTTCCCTGAT 839  
 QY 1305 GGCCTTTGAGACCCAGGCTGATCTTGAAGCACTCTCTTCTTCAAGACTC 1353  
 Db 840 GGCCTTTGAGACCCAGGCTGATCTTGAAGCACTCTCTTCTTCAAGACTC 888  
 RESULT 3  
 BM923713 1013 bp mRNA linear EST 12-MAR-2002  
 LOCUS BM923713  
 DEFINITION AGENCOUT\_6623973 NIH\_MGC\_116 Homo sapiens cDNA clone IMAGE:5759434  
 5', mRNA sequence.  
 ACCESSION BM923713  
 VERSION BM923713.1 GI:19374092  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 1013)  
 NIH-MGC http://mgc.nci.nih.gov/  
 AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished

COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cgsaps-x@mail.nih.gov  
 Tissue Procurement: Life Technologies, Inc.  
 CDNA Library Preparation: Life Technologies, Inc.  
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
 DNA Sequencing by: Agencourt Bioscience Corporation  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LNL at:  
 http://image.llnl.gov  
 plate: LLM12804 row: n column: 11  
 High quality sequence stop: 654.

FEATURES  
 source  
 1. .1013  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:5759434"  
 /lab\_host="PH10B"  
 /note="Organ: NIH\_MGC 116"  
 /note="Organ: pooled colon, kidney, stomach; Vector:  
 pcwv-sport6; Site 1: NotI; Site 2: EcoRV (destroyed); RNA  
 source anonymous pool of 3 colons, age 26 yo male, 49 yo  
 female, 71 yo male colon, 46 yo male kidney, and pool of 2  
 stomachs, 62 yo male and 70 yo female. Library is  
 oligo-dT primed and directionally cloned (EcoRV site is  
 destroyed upon cloning). Average insert size 1.4 kb,  
 insert size range 1-3 kb. Library is normalized and  
 enriched for full-length clones and was constructed by C.  
 Gruber (Invitrogen). Research Genetics tracking code  
 023. Note: this is a NIH MGC Library."

BASE COUNT 176 a 338 c 307 g 192 t  
 ORIGIN

Query Match 50.3%; Score 839.8; DB 12; Length 1013;  
 Best Local Similarity 95.6%; Pred. No. 6,4e-167;  
 Matches 894; Conservative 2; Mismatches 34; Indels 5; Gaps 3;

QY 97 GATTTCGGATCCTTCTCTGAGCCCAATGCTGGGCTCTGACACCAACCCCAAG 156  
 Db 1 GATTTCGGATCCTTCTCTGAGCCCAATGCTGGGCTCTGACACCAACCCCAAG 60  
 QY 157 GCTTCCTACCTTGCCTGAGTGTGCTGAGGAGGCTGCTGCTGCTGCTGCTGCT 216  
 Db 61 GCTTCCTACCTTGCCTGAGTGTGCTGAGGAGGCTGCTGCTGCTGCTGCTGCT 120  
 QY 217 GAAAGGGGCTCTGAGGCTCTGAGGCTCTGAGGCTCTGAGGCTCTGAGGCTCT 276  
 Db 121 GAAAGGGGCTCTGAGGCTCTGAGGCTCTGAGGCTCTGAGGCTCTGAGGCTCT 180  
 QY 277 ACTCGGCTGGGAGACAGAGCGGAGGAGGCTGCTGCTGCTGCTGCTGCTGCTGCT 336  
 Db 181 ACTCGGCTGGGAGACAGAGCGGAGGAGGCTGCTGCTGCTGCTGCTGCTGCTGCT 240  
 QY 337 ACGCATCAGAGTGGAGAGTGCAGTGCCTGCTGAGTGCCTGCTGAGTGCAGTGC 396  
 Db 241 ACGCATCAGAGTGGAGAGTGCAGTGCCTGCTGAGTGCCTGCTGAGTGCAGTGC 300  
 QY 397 CTATGAAGGCTCATGTGTGTGTGCTCTCTGCTGCTGCTGCTGCTGCTGCTGCT 456  
 Db 301 CTATGAAGGCTCATGTGTGTGTGCTCTCTGCTGCTGCTGCTGCTGCTGCTGCT 360  
 QY 457 TGTCTACTGCTTCCCAAGGAGACCAAGAGGCTGAGTGCCTGAGTGCCTGAGTGC 516  
 Db 361 TGTCTACTGCTTCCCAAGGAGACCAAGAGGCTGAGTGCCTGAGTGCCTGAGTGC 420  
 QY 517 CCAAGTGAATCTCTAATCCGAGAGCGCAAGGTGACCACTGGAAGACATATCCCA 576  
 Db 421 CCAAGTGAATCTCTAATCCGAGAGCGCAAGGTGACCACTGGAAGACATATCCCA 480  
 QY 577 CCCAGCTACTCTTCAGAGAGGCTCCCAAGGCGACATTTGACTCTCTCACTCAGCAGAC 636  
 Db 481 CCCAGCTACTCTTCAGAGAGGCTCCCAAGGCGACATTTGACTCTCTCACTCAGCAGAC 540

Query Match 49.0%; Score 816.8; DB 14; Length 865;  
 Best Local Similarity 99.2%; Pred. No. 4,5e-162;  
 Matches 829; Conservative 2; Mismatches 4; Indels 1; Gaps 1;

QY 637 CATCATTCTCCCGCTACATCCGACCCCATCTGCTCCCTGACAGCCACCTCTCC 696  
 DB 541 CATCATTCTCCCGCTACATCCGACCCCATCTGCTCCCTGACAGCCACCTCTCC 600  
 QY 697 CAACGGCTTCACATGCACTGTCACTGCTGAGGCTCATGTGAGCCCTCACTGAGCTCT 756  
 DB 601 CAACGGCTTCACATGCACTGTCACTGCTGAGGCTCATGTGAGCCCTCACTGAGCTCT 660  
 QY 757 GAGGCCCAAGCCACTGAGAGCACTCGAGGCTCTGATGATGATGATGATGATGATGAT 816  
 DB 661 GAGGCCCAAGCCACTGAGAGCACTCGAGGCTCTGATGATGATGATGATGATGATGAT 719  
 QY 817 GCCTGTACAAATGACAGCCAGCCTGAGAGCCGCACTTGTTCAGAGAGAGAGAGAGAG 876  
 DB 720 GCCTGTACAAATGACAGCCAGCCTGAGAGCCGCACTTGTTCAGAGAGAGAGAGAGAG 779  
 QY 877 GTGCTGTATATGTGAG 934  
 DB 780 GTGCTGTATATGTGAG 839  
 QY 935 CTCCTGCTGTGAG 992  
 DB 840 CTCCTGCTGTGAG 899  
 QY 993 TGTGAG 1027  
 DB 900 GGG 934

RESULT 4  
 LOCUS CD389322 865 bp mRNA linear EST 30-MAY-2003  
 DEFINITION ABEHCOURT\_14302872 NIH\_MGC\_173 Homo sapiens cDNA 5', mRNA sequence.  
 ACCESSION CD389322  
 VERSION CD389322.1 GI:31227998  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 865)  
 NIH-MGC http://mgi.nci.nih.gov/  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished  
 COMMENT Contact: Daniela S. Gerhard, Ph. D.  
 Office of Cancer Genomics / NIH  
 National Cancer Institute / NIH  
 Bldg. 31 Rm10A07 Bethesda, MD 20892  
 Email: cga@dc-remail.nih.gov  
 Tissue Procurement: Dr. Jamie Thompson, University of WI  
 cDNA Library Preparation: Gina Zastrow-Hayes  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
 DNA Sequencing by: Agencourt Bioscience Corporation  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LNL at:  
 http://image.llnl.gov  
 Plate: NDKM79 row: f column: 16  
 High quality sequence start: 12  
 High quality sequence stop: 651.  
 Location/Qualifiers

FEATURES  
 source  
 1. 865  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /tissue\_type="Trophoblast"  
 /lab\_host="DH10B TONa"  
 /clone\_lib="NIH MGC 173"  
 /note="Vector: pDONR201; Site 1: attP; Site 2: attP1;  
 LIBR PRIMING - oligo dt; METHOD - full-length enriched;  
 LIBR PROVIDER - Bradfield"

BASE COUNT 153 a 309 c 235 g 168 t  
 ORIGIN

QY 3 CACTGCTCCCTTCCGATCTCCAGAGGCTCTCCCTGAGGCGGCTCCCTGATTAAGGC 62  
 DB 22 CACTGCTCCCTTCCGATCTCCAGAGGCTCTCCCTGAGGCGGCTCCCTGATTAAGGC 81  
 QY 63 CAGCTTTGAGACATTCTGCTCCCTTCCAGCCGAGATTTGGAGATCTTCCCTGAGGC 122  
 DB 82 CAGCTTTGAGACATTCTGCTCCCTTCCAGCCGAGATTTGGAGATCTTCCCTGAGGC 141  
 QY 123 AACATGTGAGTCTGCTTCCAGACCAACCAAGCTTCTTACCTTCCCTGCTGAGATC 182  
 DB 142 AACATGTGAGTCTGCTTCCAGACCAACCAAGCTTCTTACCTTCCCTGCTGAGATC 201  
 QY 183 TGCCCAAGAGGCGCTTGTCTGAGGCAATGAGGAGAGAGAGAGAGAGAGAGAGAG 242  
 DB 202 TGCCCAAGAGGCGCTTGTCTGAGGCAATGAGGAGAGAGAGAGAGAGAGAGAGAG 261  
 QY 243 TGGGAGCTGTGAGCACTTCTGCTTATCTTGAATTAATCCGATGAGAGAGAGAGAG 302  
 DB 262 TGGGAGCTGTGAGCACTTCTGCTTATCTTGAATTAATCCGATGAGAGAGAGAGAG 321  
 QY 303 GGGCAGAGCTTCTGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 362  
 DB 322 GGGCAGAGCTTCTGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 381  
 QY 363 TGGGAGCTGTGAGCACTTCTGCTTATCTTGAATTAATCCGATGAGAGAGAGAGAG 422  
 DB 382 TGGGAGCTGTGAGCACTTCTGCTTATCTTGAATTAATCCGATGAGAGAGAGAGAG 441  
 QY 423 GCTCTCTGCTGTGAG 482  
 DB 442 GCTCTCTGCTGTGAG 501  
 QY 483 ACAAGAGAGCTTATGAGTCAAGCTGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 542  
 DB 502 ACAAGAGAGCTTATGAGTCAAGCTGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 561  
 QY 543 CCAAGGTGAG 602  
 DB 562 CCAAGGTGAG 621  
 QY 603 AGGAGCAGATTGAGCTCTTCACTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 662  
 DB 622 AGGAGCAGATTGAGCTCTTCACTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 681  
 QY 663 CCACTGCTCTCTGAG 722  
 DB 682 CCACTGCTCTCTGAG 741  
 QY 723 GCTGAGGATCATGTGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 782  
 DB 742 GCTGAGGATCATGTGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 801  
 QY 783 AGGAGCCTGTGATGAGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 838  
 DB 802 AGGAGCCTGTGATGAGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 856

RESULT 5  
 LOCUS B0690230 922 bp mRNA linear EST 15-JUL-2002  
 DEFINITION ABEHCOURT\_8343359 NIH\_MGC\_110 Homo sapiens cDNA clone IMAGE:6250272  
 5', mRNA sequence.  
 ACCESSION B0690230  
 VERSION B0690230.1 GI:2181546  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 922)  
 AUTHORS NIH-MGC <http://mgi.nci.nih.gov/>.  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)  
 Tissue Procurement: ATCC  
 cDNA Library Preparation: Rubin Laboratory  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
 DNA Sequencing by: Agencourt Bioscience Corporation  
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: <http://image.llnl.gov>  
 Plate: LNCM2393 row: b column: 01  
 High quality sequence stop: 597.  
 Location/Qualifiers

FEATURES  
 source 1..922  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:6250272"  
 /issue\_type="ductal carcinoma, cell line"  
 /lab\_host="DH10B (phage-resistant)"  
 /clone\_1ib="NIH\_MGC\_110"  
 /note="Organ: pancreas; Vector: pOTB7; Site\_1: XhoI; Site\_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGACGAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using Zap-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH\_MGC Library."

BASE COUNT 162 a 318 c 258 g 181 t 3 others

ORIGIN

Query Match 48.6%; Score 810.8; DB 13; Length 922;  
 Best Local Similarity 96.7%; Prid. No. 8.4e-161;  
 Matches 887; Conservative 2; Mismatches 22; Indels 6; Gaps 6;

3 CACTGCTCGCTGGATATCCAGAGGCTCCGTTGGCGCGCTCCCTGCTTAAGAGG 62  
 1 CACTGCTCGCTGGATATCCAGAGGCTCCGTTGGCGCGCTCCCTGCTTAAGAGG 60  
 63 CAGCCTTGAACATCTGCTGCTTCCAGCCCGGATTCGGATCCTTCCCTCTAGACC 122  
 61 CAGCCTTGAACATCTGCTGCTTCCAGCCCGGATTCGGATCCTTCCCTCTAGACC 120  
 123 AACATCTGGGCTCTGCTTGGACACACCCCAAGGCTTCTTACCTTGGCTGAGATC 182  
 121 AACATCTGGGCTCTGCTTGGACACACCCCAAGGCTTCTTACCTTGGCTGAGATC 180  
 183 TGGCCCAAGAGGCGCTTGTCTGCGGCGATGCGMAAGAGGCGCTTGGCGCGCTTGGCGGCAAG 242  
 181 TGGCCCAAGAGGCGCTTGTCTGCGGCGATGCGMAAGAGGCGCTTGGCGGCGCTTGGCGGCAAG 240  
 243 TGGGCGGCTGCGCATCTGCTTATCTTGAATTAATCTCGGATCGGAGACAGAGCGGAAG 302  
 241 TGGGCGGCTGCGCATCTGCTTATCTTGAATTAATCTCGGATCGGAGACAGAGCGGAAG 300  
 303 GGGCAGAACCTTCCTGCGGCTGCGGCGCGCAAGCAGCATCACAGGTGCGACAGTGACAG 362  
 301 GGGCAGAACCTTCCTGCGGCTGCGGCGCGCAAGCAGCATCACAGGTGCGACAGTGACAG 360  
 363 TGGCCGCTGAGTGGCGCTTGGCGGCGATGAGATCACTATGAAGCGCTTCAATGCTGAGTGG 422  
 361 TGGCCGCTGAGTGGCGCTTGGCGGCGATGAGATCACTATGAAGCGCTTCAATGCTGAGTGG 420  
 423 GCTCTCTCTGCTGAGAGAGTGGCTGCTGACAGCTCTCACTGCTTCCCAAGAGAGACCC 482  
 421 GCTCTCTCTGCTGAGAGAGTGGCTGCTGACAGCTCTCACTGCTTCCCAAGAGAGACCC 480  
 483 ACAAGAACCTTATGAGTCAAGCTTGGGCGCGCAGAGCTAAGCTCTTACTCGAGAGACG 542

DB 481 ACAAGAACCTTATGAGTCAAGCTTGGGCGCGCAGAGCTAAGCTCTTACTCGAGAGACG 540  
 QY 543 CCAAGGTGACACCCCTGAAGAGACATCATCCCAACCCAGCTACTCCAGAGAGGCTCCC 602  
 DB 541 CCAAGGTGACACCCCTGAAGAGACATCATCCCAACCCAGCTACTCCAGAGAGGCTCCC 600  
 QY 603 AAGGCGACATTGCACTCTCTCAACTCAGAGAGACCCATCACTTCTCCGCTAATCCGAG 662  
 DB 601 AAGGCGACATTGCACTCTCTCAACTCAGAGAGACCCATCACTTCTCCGCTAATCCGAG 660  
 QY 663 CCATCTGCTCTCCCTGAGAGCAACGCTCTTCTCCCAAGGCTCCAGCTGACTGTACTG 722  
 DB 661 CCATCTGCTCTCCCTGAGAGCAACGCTCTTCTCCCAAGGCTCCAGCTGACTGTACTG 720  
 QY 723 GCT-GGGGTGATGTGGCCCTCTCAGTGAAGCTCTGAGGCGCCCAAGCAGCTGAGAGACTC 781  
 DB 721 GCTGGGGGTGATGTGGCCCTCTCAGTGAAGCTCTGAGGCGCCCAAGCAGCTGAGAGACTC 779  
 QY 782 GAGGTGCTCTGTATGATGCTGTGAGAGCTGTGTAATGCTCTGTACAAATCGAGCGCAAGCC 841  
 DB 780 GAAGTGCTCTGTATGATGCTGTGAGAGCTGTGTAATGCTCTGTACAAATCGAGCGCAAGCC 837  
 QY 842 TGAAGAGCGCGCACTTTGTTCAAAGAGAGCATGCTGTGTCTGCTATGTGG-AGGGGGCA 900  
 DB 838 TGAAGAG-CGCACTTTGTTCAAAGAGAGCATGCTGTGTCTGCTATGTGG-AGGGGGCA 896  
 QY 901 AGGAGCGCTGCGCAGGGT 917  
 DB 897 AGGAGCGCTTCCAGAGT 913

RESULT 6  
 CD050808 909 bp mRNA linear EST 09-MAY-2003  
 LOCUS AGENCOURT\_13960165 NIH\_MGC\_173 Homo sapiens cDNA 5', mRNA sequence.  
 DEFINITION CD050808  
 ACCESSION CD050808.1 GI:30488103  
 VERSION  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

REFERENCE 1 (bases 1 to 909)  
 AUTHORS NIH-MGC <http://mgi.nci.nih.gov/>.  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)  
 Tissue Procurement: Dr. Jamie Thompson, University of WI  
 cDNA Library Preparation: Gina Zastrow-Hayes  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
 DNA Sequencing by: Agencourt Bioscience Corporation  
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: <http://image.llnl.gov>  
 Plate: NDKM33 row: f column: 15  
 High quality sequence start: 18  
 High quality sequence stop: 774.  
 Location/Qualifiers

FEATURES  
 source 1..909  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /issue\_type="Trophoblast"  
 /lab\_host="DH10B Tona"  
 /clone\_1ib="NIH\_MGC\_173"  
 /note="Vector: pDONR201; Site 1: attP2; Site 2: attP1; LIBR PRIMING - oligo dT; METHOD - full-length enriched; LIBR PROVIDER - Bradfield"

BASE COUNT 165 a 321 c 249 g 174 t

ORIGIN  
 Query Match 47.7%; Score 795.2; DB 14; Length 909;





Db	482	CTCCCTGAGCCAAAGCCTCTCTTCCCAAGGGCTCCACTGTGACTGTGCTGGGGT	541
QY	731	CATGAGGCCCCCTAGTAGGCTCCTGAGGCCAAGCCACTGACACCAACTGAGGTGCT	790
Db	542	CATGTGGCCCCCTCAGTAGGCTCTTGAGGCCAAGCCACTGACCACTGACGTGCT	601
QY	791	CTGATCAGTCGTGAGACGTGTGAACCTGCTGTACAAACATCGACGCCAAGCTGAGAGCC	850
Db	602	CTGATCAGTCGTGAGACGT - GTAACCTGCTGTACAAACATCGACGCCAAGCTGAGAGACC	660
QY	851	GCACTTTGTCCAAAGGACATGTGTGTGTGCTGTGCTATGTGAGGGGGGCMAAGACGCTG	910
Db	661	GCACTTTGTCCAAAGGACATGTGTGTGTGCTGTGCTATGTGAGGGGGGCMAAGACGCTT	720
QY	911	CCAGGGTGACTCTGTGGGGGCCACTCTCC - TGCCCTGTGAGAGGATCTGTGATACCTGACGG	969
Db	721	GCCAGGTGACTCTGTGGGGGCCACTCTCTTGCCCTGTGAGAGGGGCTCTGTGATCTGACGG	780
QY	970	GCATTGTGAGCTGGGGAGAGATCCTGTGGGGCCCCGACAGGCTGTGTGTGATCACTTGG	1029
Db	781	GGCTGTGTGAGCTGGGGACATGCTCTGTGGGGCCCCG - ACAGGCTGTGTGTGTGATCACTTGG	839
QY	1030	CTTCACGTAATGCTCCTCTGATC 1052	
Db	840	-CTTCACGTAATGCTCCTCTGTGATC 861	

RESULT 8	LOCUS	DEFINITION
CB995955	CB995955	815 bp mRNA linear EST 01-MAY-2003 ABENCCOURT 13511642 NIH MGC_148 Homo sapiens cDNA clone IMAGE:30330676 5', mRNA sequence.

ACCESSION	CB995955
VERSION	CB995955.1
KEYWORDS	EST.
SOURCE	Homo sapiens (human)

ORGANISM Homo sapiens  
Eukaryote; Metazoa; Chordata; Craniata; Vertebrates; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 815)  
NIH-MGC <http://mgc.ncl.nih.gov/>.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished  
Contact: Robert Strausberg, Ph.D.

Email: [cgabbs-r@mail.nih.gov](mailto:cgabbs-r@mail.nih.gov)  
Tissue Procurement: Dr. Stefan Hansson  
cDNA Library Preparation: Michael J. Brownstein (NHRI) with help  
and advice from Piero Carninci (RIKEN)  
cDNA library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNL at:  
<http://image.llnl.gov>  
Plate: NDAM34 row: 0 column: 05  
High quality sequence stop: 615.

```
FEATURES      Location/Qualifiers
source        1. .815
```

```

/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:30330676"
/tissue_type="pre-eclamptic placenta"
/lab_host="DH10B Tona"
/clone_id="NIM MGC 148"
/note="Organ: Placenta; Vector: pBluescriptR; Site 1:
all-XhoI; Site 2: BamH; Library is oligo-dT primed and
directionally cloned using primer
5'-TTTTTTTTTTTTTTTIVN-3', size-selected for average insert
size 2.3 kb and normalized to R0N 5. This is a primary
library enriched for full-length clones and constructed
using the Cap-trapper method (Garninci, in preparation).
Library constructed by M. Brownstein (NIMH/NHGRI).

```

National Institutes of Health). Note: this is a NIH_MGC Library."				
BASE COUNT	139 a	228 g	155 t	1 others
ORIGIN	292 c			

```
Query Match      45.7%;  Score 762.2;  DB 14;  Length 815;
Best Local Similarity 99.6%;  Pred. No. 1.5e-150;
Matches 772;  Conservative 2;  Mismatches 0;  Indels 1;  Gaps 1
```

[illegible]

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
1 (bases 1 to 1057)  
Li, W.B., Gruber, C., Jessee, J. and Polayes, D.  
Full-length cDNA libraries and normalization  
Unpublished  
Contact: Genoscope  
Genoscope - Centre National de Sequencage  
BP 191 91006 Evry cedex - France  
Email: segre@genoscope.cns.fr; Web : www.genoscope.cns.fr  
Library was constructed by Life Technologies, a division of  
Invitrogen. This sequence belongs to sequence cluster 7994.f For  
more information about this cluster, see  
http://www.genoscope.cns.fr/  
cgi-bin/cluster.cgi?seq=CS1A10102B01Q1&cluster=7994.f. Contact :  
Feng Liang Email : fliang@life.techn.com URL :  
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600  
Faraday Avenue Genoscope sequence ID : CS1A10102B01Q1P1.

FEATURES  
Source  
1.1057  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="CSOD1037YB01"  
/issue\_type="PLACENTA COT 25-NORMALIZED"  
/clone\_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"  
/note="1st strand cDNA was primed with a NotI-oligo (dT)  
primer. Five prime end enriched, double-strand cDNA was  
digested with Not I and cloned into the Not I and EcoR V  
sites of the pCMVSPORT 6 vector. Library was normalized."  
Location/Qualifiers

BASE COUNT  
228 a 274 c 338 g 206 t 11 others

Query Match 45.2%; Score 753.2; DB 13; Length 1057;  
Best Local Similarity 97.9%; Pred. No. 1.2e-148;  
Matches 783; Conservative 1; Mismatches 14; Indels 2; Gaps 2;

11 GCTTCGATACCTCAGAGCCCTCCCTGCGCGCGCTCCCTGCTTAGAGCCAGCCTTG 70  
11 GCTTCGATACCTCAGAGCCCTCCCTGCGCGCGCTCCCTGCTTAGAGCCAGCCTTG 810  
869 GCTTCGATACCTCAGAGCCCTCCCTGCGCGCGCTCCCTGCTTAGAGCCAGCCTTG 810  
71 GACACTTGTGCCCCCTTTCAGCCCGGATCTGGGATCTTCCCTGAGCCAACTTG 130  
809 GACACTTGTGCCCCCTTTCAGCCCGGATCTGGGATCTTCCCTGAGCCAACTTG 750  
131 GGTCTGCTCTGACACCAACCCAGCTTCTAAGCTTGGTGGCTGGAGTCTGCCAG 190  
749 GGTCTGCTCTGACACCAACCCAGCTTCTAAGCTTGGTGGCTGGAGTCTGCCAG 690  
191 GGGCCCTTGTCTGCGGACATGCGGAGAGGGGCTCTGGGGCTGGGAGCTGGGGCT 250  
689 GGGCCCTTGTCTGCGGACATGCGGAGAGGGGCTCTGGGGCTGGGAGCTGGGGCT 631  
251 GTGGCCATTCGCTATCTTGTGATTAAGCTTCTGCGGAGCAGAGCGGAGGCGAGAA 310  
630 GTGGCCATTCGCTATCTTGTGATTAAGCTTCTGCGGAGCAGAGCGGAGGCGAGAA 571  
311 GCTTCTGCGGTGTGGCCCCCAAGCAGATCAGAGTGGCAGAGTGGCGGCT 370  
570 GCTTCTGCGGTGTGGCCCCCAAGCAGATCAGAGTGGCAGAGTGGCGGCT 512  
371 CAGTGGCCCTGGCAGGTCAGATCACTATGAAGCGCTCATGTGTGTGGTGGCTCTC 430  
511 CAGTGGCCCTGGCAGGTCAGATCACTATGAAGCGCTCATGTGTGTGGTGGCTCTC 452  
431 GTGTCTGAGAGGAGGAGTGTGCTGAGCTGCTCTCCCGAGCAGACCAAGAGAA 490  
451 GTGTCTGAGAGGAGGAGTGTGCTGAGCTGCTCTCCCGAGCAGACCAAGAGAA 392  
491 GCTATGAGGAGTCAAGCTGGGGGCCCAAGCTAGACTCTTCCAGAGAGCGCAAGGTC 550  
391 GCTATGAGGAGTCAAGCTGGGGGCCCAAGCTAGACTCTTCCAGAGAGCGCAAGGTC 332  
551 AGCACTGAGAGGAGTCAATCCCGACCCAGCTAAGCTTCCAGAGAGGCTCCCAAGGCGAC 610

Db  
331 AGCACTGAGAGGAGTCAATCCCGACCCAGCTAAGCTTCCAGAGAGGCTCCAGGCGAC 272  
Qy  
611 ATTGACTCTCTCCACTAGAGAGCCCTTTCAGCTTCCCGCTCATCCGCGCCATCTGC 670  
Db  
271 ATTGACTCTCTCCACTAGAGAGCCCTTTCAGCTTCCCGCTCATCCGCGCCATCTGC 212  
Qy  
671 CTCCCTGAGCAGCAGGCTCTCTCCCGAGCGGCTCCACTGCTCACTGGTGGGT 730  
Db  
211 CTCCCTGAGCAGCAGGCTCTCTCCCGAGCGGCTCCACTGCTCACTGGTGGGT 152  
Qy  
731 CATGTGGCCCTCTAGTGGCTCTCTGAGCGCCCAAGCCACTGAGCACTGAGGTGCT 790  
Db  
151 CATGTGGCCCTCTAGTGGCTCTCTGAGCGCCCAAGCCACTGAGCACTGAGGTGCT 92  
Qy  
791 CTGATCAGTCTGAGAGCTG 810  
Db  
91 CTGATCAGTCTGAGAGCTG 72

RESULT 10  
AL575026/c 1088 bp mRNA linear EST 01-JUN-2003  
LOCUS  
DEFINITION  
AL575026 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA  
clone CSOD1064YF02 3-PRIME, mRNA sequence.  
AL575026  
AL575026.2 GI:31313335  
VERSION  
KEYWORDS  
EST.  
SOURCE  
Homo sapiens (human)  
ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
1 (bases 1 to 1088)  
Li, W.B., Gruber, C., Jessee, J. and Polayes, D.  
Full-length cDNA libraries and normalization  
Unpublished  
On Feb 16, 2001 this sequence version replaced gi:12935789.  
Contact: Genoscope  
Genoscope - Centre National de Sequencage  
BP 191 91006 Evry cedex - France  
Email: segre@genoscope.cns.fr; Web : www.genoscope.cns.fr  
Library was constructed by Life Technologies, a division of  
Invitrogen. This sequence belongs to sequence cluster 7896.r For  
more information about this cluster, see  
http://www.genoscope.cns.fr/  
cgi-bin/cluster.cgi?seq=CSOD1064DC01NP1&cluster=7896.r. Contact :  
Feng Liang Email : fliang@life.techn.com URL :  
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600  
Faraday Avenue Genoscope sequence ID : CSOD1064DC01NP1.

FEATURES  
Source  
1.1088  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="CSOD1064YF02"  
/issue\_type="PLACENTA COT 25-NORMALIZED"  
/clone\_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"  
/note="1st strand cDNA was primed with a NotI-oligo (dT)  
primer. Five prime end enriched, double-strand cDNA was  
digested with Not I and cloned into the Not I and EcoR V  
sites of the pCMVSPORT 6 vector. Library was normalized."  
Location/Qualifiers

BASE COUNT  
234 a 312 c 349 g 177 t 16 others

Query Match 45.0%; Score 751.4; DB 9; Length 1088;  
Best Local Similarity 93.8%; Pred. No. 3e-148;  
Matches 891; Conservative 8; Mismatches 37; Indels 14; Gaps 11;

Qy  
713 ACTGTCACTGGCTGAGTGTGCGCCCTCTAGTGAAGCTCTGAGCGCCAGCACTG 772  
Db  
1079 ACTGTCACTGGCTGAGTGTGCGCCCTCTAGTGAAGCTCTGAGCGCCAGCACTG 1023  
Qy  
773 CAGCACTGAGGAGTCTGATCAGTCTGAGAGCGTGAAGTGAAGTGAAGTGAAGTGA 832



```

Db      1022 CACCAACTCGA-GTGGCTGTGTCAGTCGAGACGT-GTAACTGGCTGTACAAATCGA 965
Qy      833 CGCAAGCTGAGAGAGCCGACATTTGTCCAAAGAGACATGATGTGTGTGGATGTGG 892
Db      964 CGCAAGCTGAGAGAGCCGACATTTGTCCAAAGAGACATGATGTGTGTGGATGTGG 905
Qy      893 GGGGGGCAAGAGAGCGCTGCGAGGGTGAATCTGTGGGGGCCACTCTCTGCTGTGAGAG 952
Db      904 -GGGGGCAAGAGAGCGCTGCGAGGGTGAATCT-GGGGGCCACTCTCTGCTGTGAGAG 847
Qy      953 TCTCTGTGTACTGAAGGGGATTTGTAGCTGTGGGAGATGCTGTGGGGCCCGCAAGGCC 1012
Db      846 TCTCTGTGTACTGAAGGGGATTTGTAGCTGTGGGAGATGCTGTGGGGCCCGCAAGGCC 787
Qy      1013 TGGTGTGTACTGAAGGGGATTTGTAGCTGTGGGAGATGCTGTGGGGCCCGCAAGGCC 1072
Db      786 TGGTGTGTACTGAAGGGGATTTGTAGCTGTGGGAGATGCTGTGGGGCCCGCAAGGCC 727
Qy      1073 CCAAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1132
Db      726 CCAAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 668
Qy      1133 CCAAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1192
Db      667 CCAAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 608
Qy      1193 TCTGGGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1252
Db      607 TCTGGGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 548
Qy      1253 TTCCAGAGATGATGATCACTCACTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1312
Db      547 TTCCAGAGATGATGATCACTCACTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 488
Qy      1313 GAGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1372
Db      487 GAGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 428
Qy      1372 ATCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1432
Db      427 ATCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 368
Qy      1433 GAGTTTACTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1492
Db      367 GAGTTTACTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 308
Qy      1493 CATTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1550
Db      307 CATTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 249
Qy      1551 TCCCACTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1610
Db      248 TCCCACTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 192
Qy      1611 TTTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1660
Db      191 CTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 142

```

```

RESULT 11
LOCUS      BX386211 1201 bp mRNA linear EST 08-MAY-2003
DEFINITION BX386211 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA
            clone CS0D1050YB10 5-PRIME, mRNA sequence.
ACCESSION  BX386211.1 GI:30449421
VERSION     BX386211.1
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 1201)

```

```

AUTHORS    Li W.B., Gruber,C., Jesse,J. and Polayes,D.
TITLE      Full-length cDNA libraries and normalization
JOURNAL    Unpublished
COMMENT    Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 Evry cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            Library was constructed by Life Technologies, a division of
            Invitrogen. This sequence belongs to sequence cluster 7896.r For
            more information about this cluster, see
            http://www.genoscope.cns.fr/
            cgi-bin/cluster.cgi?seq=CS1A10132D06QPLcluster=7896.r. Contact :
            Feng Liang Email : fliang@life.techn.com URL :
            http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
            Faraday Avenue Genoscope sequence ID : CS1A10132D06QPL.

FEATURES
  source
    1..1201
    /organism="Homo sapiens"
    /mol_type="mRNA"
    /db_xref="taxon:9606"
    /clone="CS0D1050YB10"
    /tissue_type="PLACENTA COT 25-NORMALIZED"
    /clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"
    /note="1st strand cDNA was primed with a NotI-oligo(dT)
    primer. Five prime end enriched, double-strand cDNA was
    digested with Not I and cloned into the Not I and BclR V
    sites of the pCMVSPORT 6 vector. Library was normalized."

BASE COUNT  212 a 361 c 309 g 223 t
ORIGIN
Query Match 44.6%; Score 743.6; DB 13; Length 1201;
Best Local Similarity 83.3%; Pred. No. 1.3e-146;
Matches 907; Conservative 58; Mismatches 100; Indels 24; Gaps 12;

Qy      43 CGGCTCCGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 102
Db      42 CCGGATCTGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 101
Qy      103 GGGATCTTCCCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 162
Db      102 GGGATCTTCCCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 161
Qy      163 TACCTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 222
Db      162 TACCTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 221
Qy      223 GGTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 282
Db      222 GGTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 281
Qy      283 GTGGGGGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 342
Db      282 GTGGGGGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 341
Qy      343 CACAGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 402
Db      342 CACAGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 401
Qy      403 AGGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 461
Db      402 AGGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 461
Qy      462 ACTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 516
Db      462 ACTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 521
Qy      517 --CCAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 573
Db      522 CMCNGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 581
Qy      574 CCACCCAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 632
Db      582 CCNCCCGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 641

```

```

QY 633 G-ACCATCATCTTCTCCGGCTACATCCGGCCCATCTGCTCCCTGAGCAACGGCTCC 691
DB 642 GAACCAATACCTTTCCTCCGTACATMGCCCATCTGCTCCCTGAGCAACGGCTCC 701
QY 692 TTCCCAACGGGCTCTCACTGATCTGCTGAGGATCATATGAGCCCTTCAGTAGC 751
DB 702 TTGCCCAACGGAAACATCTMTGTCAMMGSTGGGCTCAATATGCAACCCAGTAGAG 761
QY 752 CTCCTGACGCCCAAGCACTGACGAACTTGAGAGTCTCTGATCATGCTGAGAGCTGG 811
DB 762 CTCCTGACGCCCAAGCACTGAGAGTCTCTGATCATGAGAGTCTGAGAGTCTGAG 820
QY 812 TAACCTGCTGATCAATGACAGCCCAAGCTGAGAGTCTCTGATCATGCTGAGAGCT 871
DB 821 TAACCTGCTGATCAATGACAGCCCAAGCTGAGAGTCTCTGATCATGAGAGTCTGAG 880
QY 872 GGTGTGCTGCTGATGATGAGAGGAGGAGCAAGAGAGCTGCTGAGAGTCTGAGAGCT 931
DB 881 GGTGTGCTGCTGATGATGAGAGGAGGAGCAAGAGAGCTGCTGAGAGTCTGAGAGCT 940
QY 932 ACTCTCTGCTGCTGAGAGGCTCTGATCACTGAGAGGAGTCTGAGAGTCTGAGAGT 991
DB 941 ACTCTCTGCTGCTGAGAGGCTCTGATCACTGAGAGGAGTCTGAGAGTCTGAGAGT 997
QY 992 CTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1051
DB 998 CTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1055
QY 1052 CCAAGCAAGGATGACAGAGTCTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1111
DB 1056 CCAAGCAAGGATGACAGAGTCTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1108
QY 1112 CGACAGCAA 1120
DB 1109 GMAVCACHA 1117

RESULT 12
BX098362 768 bp mRNA linear EST 04-FEB-2003
LOCUS BX098362 Soares breast 2NBH8st Homo sapiens cDNA clone
DEFINITION IMAGE:153925, mRNA sequence.
ACCESSION BX098362
VERSION BX098362.1 GI:27829367
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Eumleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 768)
Biert,L., Heil,O., Hennig,S., Neubert,P., Patsch,E., Peters,M.,
Radelof,U., Schneider,D. and Korn,B.
Human Unigeneset - RZPD3
Unpublished
Contact: Ina Rolfs
RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH
Im Neuenheimer Feld 580, D-69120 Heidelberg, Germany
RZPD; IMAGE:153925.
RZPDLIB; I.M.A.G.E. cDNA Clone Collection;
Human Unigeneset - RZPD3 (RZPDLIB No.972)
http://www.rzpd.de/ClonesCards/cgi-
bin/showlib.pl.cgi?responseFileNo=972 Contact: Ina Rolfs
RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH
Heubnerweg 6, D-14059 Berlin, Germany
Tel: +49 30 32639 101
Fax: +49 30 32639 111
www.rzpd.de
This clone is available royalty-free from RZPD.
Contact RZPD (clone@rzpd.de) for further information. Seq primer:
M3r, Primer sequence: TTTCACACAGGAAACAGCTATGAC.
FEATURES
SOURCE 1..768
Location/Qualifiers

```

```

/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:153925"
/sex="Female"
/dev_stage="adult"
/lab_host="DH10B (ampicillin resistant)"
/clone_lib="Soares breast 2NBH8st"
/notes="Organ: breast; Vector: pTZ19 (Pharmacia) with a
modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5']
TGTTACCATCTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG
(Pharmacia), digested with Not I and Eco RI and ligated into the Not I
and Eco RI sites of a modified pTZ19 vector (Pharmacia).
Library went through one round of normalization to a Cot =
230. Library constructed by Bento Soares and M.Fatima
Bonaldo."
BASE COUNT 129 a 267 c 212 g 160 t
ORIGIN
Query Match 4.6%; Score 743.4; DB 13; Length 768;
Best Local Similarity 99.7%; Pred. No. 1.4e-146;
Matches 755; Conservative 0; Mismatches 1; Indels 1; Gaps 1;
629 AGGAGACCATCATCTTCCGCTACATCCGAGCCATCTGCTCCCTGACCAAGCC 688
DB 13 AGGAGACCATCATCTTCCGCTACATCCGAGCCATCTGCTCCCTGACCAAGCC 72
QY 689 TCCTTCCCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 748
DB 73 TCCTTCCCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 132
QY 749 AGCTCTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 808
DB 133 AGCTCTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 192
QY 809 TGCTTAACCTGCTGATCAATGACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 868
DB 193 T-GTAACTGCTGATCAATGACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 251
QY 869 CATGAGTGTGCTGAGTGTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 928
DB 252 CATGAGTGTGCTGAGTGTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 311
QY 929 CCACTCTCTGCTGCTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 988
DB 312 CCACTCTCTGCTGCTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 371
QY 989 TGCTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1048
DB 372 TGCTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 431
QY 1049 GATCCAAAGCAAGTGAAGAGAGTCAAGAGTCAAGAGTCAAGAGTCAAGAGTCAAGAG 1108
DB 432 GATCCAAAGCAAGTGAAGAGTCAAGAGTCAAGAGTCAAGAGTCAAGAGTCAAGAGTCA 491
QY 1109 GCCGACAGAGAGTCTGTGAGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1168
DB 492 GCCGACAGAGAGTCTGTGAGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 551
QY 1169 GGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1228
DB 552 GGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 611
QY 1229 CAGGAGCACTGAGTGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1288
DB 612 CAGGAGCACTGAGTGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 671
QY 1289 CTGTCTCTGCTGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1348
DB 672 CTGTCTCTGCTGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 731

```

QY 1349 GACTCTCGGAGGCTGGGGCCCATCTTATCTTTG 1385  
 Db 732 GACTCTCGGAGGCTGGGGCCCATCTTATCTTTG 768

RESULT 13  
 BG482196 765 bp mRNA linear EST 21-MAR-2001  
 LOCUS 602526716F1 NIH\_MGC\_21 Homo sapiens cDNA clone IMAGE:4650463 5',  
 DEFINITION mRNA sequence.  
 ACCESSION BG482196  
 VERSION BG482196.1 GI:13414475  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 765)  
 NIH-MGC http://mgc.nci.nih.gov/  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 Unpublished  
 Contact: Robert Strausberg, Ph.D.  
 Email: cgabbs-remail.nih.gov  
 Tissue Procurement: ATCC  
 cDNA Library Preparation: Ling Hong/Rubin Laboratory  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LMNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LMNL at:  
 http://image.llnl.gov  
 Plate: LLM1431 row: p column: 04  
 High quality sequence stop: 765.  
 Location/Qualifiers  
 1..765  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:4650463"  
 /issue\_type="choriocarcinoma"  
 /lab\_host="DH10B (phage-resistant)"  
 /clone\_1lb="NIH\_MGC\_21"  
 /note="Organ: placenta; Vector: pOTB7; Site\_1: XhoI;  
 Site\_2: EcoRI; cDNA made by oligo-dT priming.  
 Directionally cloned into EcoRI/XhoI sites using the  
 following 5' adaptor: GGCAAGG(G). Site-selected >500bp  
 for average insert size 1.8kb. Library constructed by  
 Ling Hong in the laboratory of Gerald M. Rubin (University  
 of California, Berkeley) using ZAP-cDNA synthesis kit  
 (Stratagene) and Superscript II RT (Life Technologies)."

BASE COUNT 127 a 276 c 213 g 149 t

Query Match 44.4%; Score 741; DB 10; Length 765;  
 Best Local Similarity 99.3%; Freq. No. 4.4e-146;  
 Matches 752; Conservative 2; Mismatches 2; Indels 1; Gaps 1;

QY 15 CGGATATCTCAGGCGCTCTCCGTTGGGCGCGCTCCCTGAGAGCCAGGCTTGGAGA 74  
 Db 2 CGGATATCTCAGGCGCTCTCCGTTGGGCGCGCTCCCTGAGAGCCAGGCTTGGAGA 61  
 QY 75 CTGTGCGCCCTTTTCAGCCCGGATCTTGAGATCTTCCCTTGAGCCAAATCTGGATC 134  
 Db 62 CTGTGCGCCCTTTTCAGCCCGGATCTTGAGATCTTCCCTTGAGCCAAATCTGGATC 121  
 QY 135 CTGCGCTTGACACCAAGCCCAAGGCTTCACTTGGGTGCTGAGCTTGGCCCAAGGAGC 194  
 Db 122 CTGCGCTTGACACCAAGCCCAAGGCTTCACTTGGGTGCTGAGCTTGGCCCAAGGAGC 181  
 QY 195 CCTTGTCTGGGCGATGCGAGAGAGGAGGCTCTGGGGCTGGGAGCTGGGGGCTGTTG 254  
 Db 182 CCTTGTCTGGGCGATGCGAGAGAGGAGGCTCTGGGGCTGGGAGCTGGGGGCTGTTG 241  
 QY 255 CCATTCTGCTTATCTTGGATTACTCGGTGCGGAGCAGAGAGCGGAAGGGGAGAGTCT 314

Db 242 CCATTCTGCTTATCTTGGATTACTCGGTGCGGAGCAGAGAGCGGAAGGGGAGAGTCT 301  
 QY 315 CCTGCGGTGTGGCCCCCAAGACACGATCATCAGAGTGGCAGAGTGCCTGCTCATG 374  
 Db 302 CCTGCGGTGTGGCCCCCAAGACACGATCATCAGAGTGGCAGAGTGCCTGCTCATG 361  
 QY 375 GGCCCTGGAGGCTCAGATCATCCTATGAGAGCGTCCATGTGTGTGGTCTCTCGTGT 434  
 Db 362 GGCCCTGGAGGCTCAGATCATCCTATGAGAGCGTCCATGTGTGTGGTCTCTCGTGT 421  
 QY 435 CTGAGCAGTGGGTGCTGTGAGTGTCTCACTGCTTCCCAAGCAGACCAAGAGAGCCT 494  
 Db 422 CTGAGCAGTGGGTGCTGTGAGTGTCTCACTGCTTCCCAAGCAGACCAAGAGAGCCT 481  
 QY 495 ATGAGGTCAAGCTGGGGGCCACACAGTGTGACTCTTCTCCAGAGACCCAGAGTACGA 554  
 Db 482 ATGAGGTCAAGCTGGGGGCCACACAGTGTGACTCTTCTCCAGAGACCCAGAGTACGA 541  
 QY 555 CCTGAGAGACATCATCCCCACACCCAGGACTACCTCCAGAGAGGCTCCAGAGGACATTTG 614  
 Db 542 CCTGAGAGACATCATCCCCACACCCAGGACTACCTCCAGAGAGGCTCCAGAGGACATTTG 601  
 QY 615 CACTCTTCAACTCAGACAGACCCATCATCCTTCCCGCTTACATCCGCCATCTGCTCC 674  
 Db 602 CACTCTTCAACTCAGACAGACCCATCATCCTTCCCGCTTACATCCGCCATCTGCTCC 661  
 QY 675 CTGACAGCAAGCCTCTTCCCAAGGCTCTCACTGCTCACTGCTGAGGCTCATG 734  
 Db 662 CTGACAGCAAGCCTCTTCCCAAGGCTCTCACTGCTCACTGCTGAGGCTCATG 720  
 QY 735 TGGCCCCCTCAGTGAAGCCTCTTGAAGCCCAAGCCACT 771  
 Db 721 TTGGCCCCCTCAGTGAAGCCTCTTGAAGCCCAAGCCACT 757

RESULT 14  
 B1768651 748 bp mRNA linear EST 25-SEP-2001  
 LOCUS 603056994F1 NIH\_MGC\_122 Homo sapiens cDNA clone IMAGE:5206439 5',  
 DEFINITION mRNA sequence.  
 ACCESSION B1768651  
 VERSION B1768651.1 GI:15760229  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 748)  
 NIH-MGC http://mgc.nci.nih.gov/  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 Unpublished  
 Contact: Robert Strausberg, Ph.D.  
 Email: cgabbs-remail.nih.gov  
 Tissue Procurement: Life Technologies, Inc.  
 cDNA Library Preparation: Life Technologies, Inc.  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LMNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LMNL at:  
 http://image.llnl.gov  
 Plate: LLM1517 row: 1 column: 24  
 High quality sequence stop: 746.  
 Location/Qualifiers  
 1..748  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:5206439"  
 /lab\_host="DH10B"  
 /note="Organ: pooled lung and spleen; Vector: pCMV-SPORT6;  
 Site\_1: NotI; Site\_2: EcoRV (destroyed); RNA source"

anonymous pool of 24 week female lung, 16 week female spleen, and 20-22 week male spleen. Library is oligo-dT primed and directionally cloned (EcoRV site is destroyed upon cloning). Average insert size 1.4 kb, insert size range 1-3 kb. Library is normalized and enriched for full-length clones and was constructed by C. Gruber (Invitrogen). Research Genetics tracking code 026. Note: this is a NIH MGC library."

BASE COUNT 125 a 270 c 208 g 145 t  
 ORIGIN

Query Match 44.3%; Score 738.4; DB 12; Length 748;  
 Best Local Similarity 99.3%; Pred. No. 1.5e-145;  
 Matches 739; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 42 GCGCTCCCTGCTTGAAGGCGCAGCTTGGACCTTGTCCCTTCCAGCCGATTC 101  
 Db 4 GCGCTCCCTGCTTGAAGGCGCAGCTTGGACCTTGTCCCTTCCAGCCGATTC 63  
 QY 102 TGGGATTCCTTCTTGAAGGCAATCTGGTCTTGTGACACACCCCAAGGCTTC 161  
 Db 64 TGGGATTCCTTCTTGAAGGCAATCTGGTCTTGTGACACACCCCAAGGCTTC 123  
 QY 162 CTACCTTGTGCTGCTGAGAGTCTGCCCCAGGGGCTTGTCTGCGCCATGCGCMAGAGG 221  
 Db 124 CTACCTTGTGCTGCTGAGAGTCTGCCCCAGGGGCTTGTCTGCGCCATGCGCMAGAGG 183  
 QY 222 GGGTCTTGGGGCTTGGGAGCTGCGGAGCTGTGGCATTCTGCTTATCTTGATTAATTC 281  
 Db 184 GGGTCTTGGGGCTTGGGAGCTGCGGAGCTGTGGCATTCTGCTTATCTTGATTAATTC 243  
 QY 282 GGTTCGGGGAAGAGAGCGGAGGCGGAGAGCTTCGCGGTGGGCCCCCAAGCAGCA 341  
 Db 244 GGTTCGGGGAAGAGAGCGGAGGCGGAGAGCTTCGCGGTGGGCCCCCAAGCAGCA 303  
 QY 342 TCACAGGTGACAGAGTGCAGTGCCTGATGAGTGGCCCTGAGAGTCACTTATG 401  
 Db 304 TCACAGGTGACAGAGTGCAGTGCCTGATGAGTGGCCCTGAGAGTCACTTATG 363  
 QY 402 AAGGCGTCACTGT 461  
 Db 364 AAGGCGTCACTGT 423  
 QY 462 AAGGCGTCACTGT 521  
 Db 424 AAGGCGTCACTGT 483  
 QY 522 TAGACTCTACTCCGAGGAGCGCAAGGTCAAGCACTGAAAGACATATCCGCCACCA 581  
 Db 484 TAGACTCTACTCCGAGGAGCGCAAGGTCAAGCACTGAAAGACATATCCGCCACCA 543  
 QY 582 GTCACCTCCAGAGAGGCTCCAGGAGCGCATTTGACATCTCTCAACTAGAGACCATCA 641  
 Db 544 GTCACCTCCAGAGAGGCTCCAGGAGCGCATTTGACATCTCTCAACTAGAGACCATCA 603  
 QY 642 CCTTCTCCCGTAAATCCGCGCCATTCGCTCCCTGAGCCAAAGCTCTTCCCAAG 701  
 Db 604 CCTTCTCCCGTAAATCCGCGCCATTCGCTCCCTGAGCCAAAGCTCTTCCCAAG 663  
 QY 702 GCTCTCACTGACATGTCAGTGGGTGATGTGTGCCCCCTCAGTAGAGCTCTCTGAGC 761  
 Db 664 GCTCTCACTGACATGTCAGTGGGTGATGTGTGCCCCCTCAGTAGAGCTCTCTGAGC 723  
 QY 762 CCAAGCACTGACAGCACTGAGG 785  
 Db 724 CCAAGCACTGACAGCACTGAGG 747

RESULT 15  
 CA487658 840 bp mRNA linear EST 14-NOV-2002  
 LOCUS AGENCOURT 10808583 MAPcl Homo sapiens cDNA clone IMAGE:6719162 5',  
 DEFINITION mRNA sequence.

ACCESSION CA487658  
 VERSION CA487658.1 GI:24947600  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 840)  
 NIH-MGC http://mgs.nci.nih.gov/.  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cga@bbs-r@mail.nih.gov  
 Tissue Procurement: Kristi A. Eglund, Ira Pastan  
 cDNA Library Preparation: Invitrogen Corp  
 DNA Sequencing: Agencourt Bioscience Corporation  
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:  
 http://image.llnl.gov  
 Plate: LLNL14276 row: k column: 02  
 High quality sequence stop: 668.  
 Location/Qualifiers  
 1..840  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:6719162"  
 /cell\_line="ZR-75-1, MCF7, SK-BR-3, MDA-MB-231, HTERT-HME1", LNCaP"  
 /lab\_host="BMDH10B"  
 /clone\_id="MAPCL"  
 /note="Vector: PCMV-SPORE6; Site 1: EcoRV; Site 2: Not I; Subtracted with brain, liver, lung, kidney and muscle. Directionally cloned. Priming method: oligo-dT. Average insert size: 1800 bp. Library amplification: 26,000 fold. Kristi A. Eglund, James J. Vincent, Robert Strausberg, Bungkok Lee & Ira Pastan. Discovery of new breast cancer genes encoding membrane and secreted proteins. Manuscript submitted."

BASE COUNT 143 a 271 c 246 g 180 t  
 ORIGIN

Query Match 43.9%; Score 732.4; DB 14; Length 840;  
 Best Local Similarity 98.4%; Pred. No. 2.9e-144;  
 Matches 747; Conservative 2; Mismatches 9; Indels 1; Gaps 1;

QY 761 CCCAAGCACTGACAGCACTGAGGTGCTTGTATGATGTCAGACGTGTAACTGCT 820  
 Db 1 CCCAAGCACTGACAGCACTGAGGTGCTTGTATGATGTCAGACGTGTAACTGCT 59  
 QY 821 GTACAACTGAGAGCCCAAGCTGAGAGCGGCACTTGTCAAGAGACATGAGTGTGC 880  
 Db 60 GTACAACTGAGAGCCCAAGCTGAGAGCGGCACTTGTCAAGAGACATGAGTGTGC 119  
 QY 881 TGGCTATGTGAGAGGAGCAAGAGCGCTGSCAGAGTGACTCTGAGGAGCCCACTTCTG 940  
 Db 120 TGGCTATGTGAGAGGAGCAAGAGCGCTGSCAGAGTGACTCTGAGGAGCCCACTTCTG 179  
 QY 941 CCTGTGAGAGGCTCTGTGTAAGTGAAGGCACTTGTGAGTGTGAGATGCTGTGAGGC 1000  
 Db 180 CCTGTGAGAGGCTCTGTGTAAGTGAAGGCACTTGTGAGTGTGAGATGCTGTGAGGC 239  
 QY 1001 CCGCAACAGGCTGTGTGTGACATCTGAGGCTCAGTATGAGCTCCGATCCAAAGCAA 1060  
 Db 240 CCGCAACAGGCTGTGTGTGACATCTGAGGCTCAGTATGAGCTCCGATCCAAAGCAA 299  
 QY 1061 GGTGACAGAACTCAGAGCTGTGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 1120  
 Db 300 GGTGACAGAACTCAGAGCTGTGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 359  
 QY 1121 CTTCTGTGAGAGCACTGTGAGCTTCACTGTGCCCCAGGCCAGAGGCTTGTGAGGCCAT 1180

```
Db      360 CCTCTGTGGAGGACCACTGGACCTTCAAGCTTGCCCCAAGGCCCTTGCTGAGGCCAT 419
Qy      1181 CCTTTTCGTGCTCTGGGCTTGCTGAGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTG 1240
Db      420 CCTTTTCGTGCTCTGGGCTTGCTGAGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTG 479
Qy      1241 AGCTGGCCCTACTTCCAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1300
Db      480 AGCTGGCCCTACTTCCAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 539
Qy      1301 TGATGGCTTTTGAGACCAAGGAGCTGATGAGGACCTCTCTCTCTCTCTCTCTCTCTCTCT 1360
Db      540 TGATGGCTTTTGAGACCAAGGAGCTGATGAGGACCTCTCTCTCTCTCTCTCTCTCTCTCT 599
Qy      1361 GGCTGGGAGCCCATCTTGAATCTTTGAGCCCATCTTCTGGGTGTGCTTTTGGGACCATC 1420
Db      600 GGCTGGGAGCCCATCTTGAATCTTTGAGCCCATCTTCTGGGTGTGCTTTTGGGACCATC 659
Qy      1421 ACTGAGAGTCAGAGATTTTACTGCTGTGCAATGAGCCAGAGCTCTGGGCCCTCAGCCCA 1480
Db      660 ACTGAGAGTCAGAGATTTTACTGCTGTGCAATGAGCCAGAGCTCTGGGCCCTCAGCCCA 719
Qy      1481 CCATGGACCAAGCCCATTTGGGCGAGATCTCTGGGAGTCT 1519
Db      720 CCATGGACCAAGCCCATTTGGGCGAGATCTCTGGGAGTCT 758
```

Search completed: December 15, 2003, 23:32:19  
Job time : 3752 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 15, 2003, 17:47:59 ; Search time 480 Seconds  
(without alignments)  
9380.547 Million cell updates/sec

Title: US-09-925-301-208

Perfect score: 1668  
Sequence: 1 cacactgctcgcctcgataa.....tttggaaatttgagagaa 1668

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

N.GeneSeq 190un03.\*  
1: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/NA1980.DAT.\*  
2: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/NA1981.DAT.\*  
3: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/NA1982.DAT.\*  
4: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/NA1983.DAT.\*  
5: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/NA1984.DAT.\*  
6: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/NA1985.DAT.\*  
7: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/NA1986.DAT.\*  
8: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/NA1987.DAT.\*  
9: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/NA1988.DAT.\*  
10: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/NA1989.DAT.\*  
11: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/NA1990.DAT.\*  
12: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/NA1991.DAT.\*  
13: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/NA1992.DAT.\*  
14: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/NA1993.DAT.\*  
15: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/NA1994.DAT.\*  
16: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/NA1995.DAT.\*  
17: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/NA1996.DAT.\*  
18: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/NA1997.DAT.\*  
19: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/NA1998.DAT.\*  
20: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/NA1999.DAT.\*  
21: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/NA2000.DAT.\*  
22: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/NA2001A.DAT.\*  
23: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/NA2001B.DAT.\*  
24: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/NA2002.DAT.\*  
25: /SIDSI/gcgdata/geneSeq/geneSeq-emb1/NA2003.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	1662	99.6	1668	21	AACT7814	Human cancer assoc
2	1549.4	92.9	1835	22	AAF98720	Human late stage o
3	1549.4	92.9	3382	24	ABZ35336	Human gene express
4	1537.4	92.2	1834	24	ABST76501	cDNA encoding huma
5	1537.4	92.2	1834	24	ABN95716	Gene #2214 used to
6	1537.4	92.2	1834	24	ABK12241	cDNA encoding huma
7	1537.4	92.2	1834	24	ABL67949	Ovary cancer relat
8	1537.4	92.2	1834	24	ABL68248	Kidney cancer rela

9	1537.4	92.2	1834	24	ABL68512	Kidney cancer rela
10	1475.8	88.5	1796	22	AAF98698	Human ovarian cance
11	1420.8	85.2	1733	25	ABT13936	Human breast cancer
12	1305	78.2	1726	24	ABN95392	Human NOV14, pros
13	879.2	52.7	2131	25	ABZ80059	Human AGR receptor
14	859	51.5	882	24	ABN85395	Partial Human NOV1
15	857.4	51.4	882	24	ABN85394	Partial Human NOV1
16	778.2	46.7	1142	21	AACT8796	Activation constru
17	778.2	46.7	1142	22	AAFS5268	Nucleotide sequenc
18	778.2	46.7	1169	22	AACT8795	Activation constru
19	778.2	46.7	1169	22	AAFS5267	Nucleotide sequenc
20	744.2	44.6	1179	25	ABNS56430	Mouse channel acti
21	741.8	44.5	1161	24	ABN85393	Human NOV14b, pros
22	594.2	35.6	596	22	AAI29377	Colon tumour relat
23	594.2	35.6	596	25	ABZ33563	Human colon tumour
24	220.6	13.2	1102	24	ABQ93901	Human prostatein pr
25	220.6	13.2	1102	24	ABQ93902	Human prostatein pr
26	207.2	12.4	843	24	ABQ86175	Novel human gene.
27	207.2	12.4	849	24	ABQ86176	Novel human gene.
28	207.2	12.4	944	24	ABA94396	Human prostatein-11
29	207.2	12.4	1613	22	AAO13360	Human serine prote
30	205.8	12.3	1130	22	AAO13366	Recombinant human
31	205.6	12.3	1606	24	ABSG6730	DNA encoding prima
32	197.4	11.8	1157	22	AAI60330	Human polynucleoti
33	195.8	11.7	1130	22	AAFT7000	Fusion gene of pro
34	192.8	11.6	873	24	ABX11797	DNA encoding human
35	192.8	11.6	980	19	AAV59136	Nucleotide sequenc
36	192.8	11.6	1110	22	AAFT6994	Human protease T c
37	192.8	11.6	1129	22	AAST1354	Human cDNA sequenc
38	192.8	11.6	1129	25	ACA03713	DNA encoding huma
39	192.8	11.6	1129	25	ACA04134	Human cDNA encodin
40	192.8	11.6	1129	25	ABX89251	DNA encoding novel
41	192.8	11.6	1151	22	AAO8286	Human secreted pro
42	192.8	11.6	1703	21	AAZ52473	Human clone 3376404
43	191.2	11.5	1212	22	AAI58544	Human polynucleoti
44	184.6	11.1	786	24	ABA94394	Human prostatein-11
45	184.6	11.1	2122	23	AAST4051	DNA encoding novel

#### ALIGNMENTS

RESULT 1	AACT7814	AACT7814 standard; cDNA, 1668 BP.
ID	AACT7814;	
AC	AACT7814;	
XX		
DT	08-FEB-2001 (first entry)	
XX		
DE	Human cancer associated gene sequence SEQ ID NO:208.	
XX		
KW	Human; cancer associated gene; cancer antigen; detection; cancer;	
KW	diagnosis; cytostatic; proliferative; vulnary; immunomodulator;	
KW	antidiabetic; antiaesthetic; antirheumatic; antiarthritic; antiviral;	
KW	antiinflammatory; antihypoid; antiallergic; antibacterial; cardiant;	
KW	dermatological; neuroprotective; thrombolytic; coagulant; noctropic;	
KW	vasotrophic; antipsoriatic; antiangiogenic; gene therapy; inflammation;	
KW	immune disorder; haematopoietic cell disorder; autoimmune disorder;	
KW	allergic reaction; graft versus host disease; organ rejection;	
KW	haemostatic; thrombolytic; cardiovascular disorder; infection;	
KW	neurological disease; drug screening; ss.	
XX		
OS	Homo sapiens.	
XX		
PN	WO200055350-A1.	
XX		
PD	21-SEP-2000.	
XX		
PF	08-MAR-2000; 2000WO-US05882.	
XX		
PR	12-MAR-1999; 99US-0124270.	
XX		







Db	1221	TGGCTCTGGGGCTCTCTCTCTCCCAATGACTCAGAGAGACTGAGACTGAGCTGAGCTTACTTCCAGGA	1288
Qy	1261	TGAGATGCATCACTCAAGGACAGGAGCCTGATCTTCCCTTGATGAGCTTTGGACCCAGG	1320
Db	1281	TGAGATGATCACTCAAGGACAGGAGCCTGATCTTCCCTTGATGAGCTTTGGACCCAGG	1340
Qy	1321	GCTGACTTGGAGCCACTCTCTCTCTTCAAGACTCTGCGGGAGGCTGGGGCCCCATCTTGAT	1380
Db	1341	GCTGACTTGGAGCCACTCTCTCTCTTCAAGACTCTGCGGGAGGCTGGGGCCCCATCTTGAT	1400
Qy	1381	CTTTGAGGCCATCTTCTGGGGAGCTTTTGGGACCATCACTGAGATCGAGAGTTTAA	1440
Db	1401	CTTTGAGGCCATCTTCTGGGGAGCTTTTGGGACCATCACTGAGATCGAGAGTTTAA	1460
Qy	1441	CTGCGCTGAGCAATGGACAGAGCCTCTGAGCCCTCAGCCACCATGAGCAGAGCCATTGGS	1500
Db	1461	CTGCGCTGAGCAATGGACAGAGCCTCTGAGCCCTCAGCCACCATGAGCAGAGCCATTGGS	1520
Qy	1501	CGAGATCTCGGGAGG-TCTTGGGACCTTGGTATGAATAAGAGCCCTGGGTTTCCACCT	1558
Db	1521	CGAGATCTCGGGAGGCTCTGAGGACCTTGGTATGAATAAGAGCCCTGGGTTTCCACCT	1578
Qy	1559	GTTTCTGNAAGACTGCTTCCCGGAGCCGCGCTTCCAGACTGATGATGACACTTTTCTTTC	1618
Db	1580	GTTTCTGNAAGACTGCTTCCCGGAGCCGCGCTTCCAGACTGATGATGACACTTTTCTTTC	1638
Qy	1619	CNTTTCCTGTTGTTTGGGATGGGCACTTTTGGAGTTT	1660
Db	1637	CCTCTCCCTGTTGTTGAGGCTGGGCACTTTTGGAGTTT	1678

### RESULT 3

ID	ABZ35336	standard; cDNA; 3382 BP.
1	1	1
2	2	2
3	3	3
4	4	4
5	5	5
6	6	6
7	7	7
8	8	8
9	9	9
10	10	10
11	11	11
12	12	12
13	13	13
14	14	14
15	15	15
16	16	16
17	17	17
18	18	18
19	19	19
20	20	20
21	21	21
22	22	22
23	23	23
24	24	24
25	25	25
26	26	26
27	27	27
28	28	28
29	29	29
30	30	30
31	31	31
32	32	32
33	33	33
34	34	34
35	35	35
36	36	36
37	37	37
38	38	38
39	39	39
40	40	40
41	41	41
42	42	42
43	43	43
44	44	44
45	45	45
46	46	46
47	47	47
48	48	48
49	49	49
50	50	50
51	51	51
52	52	52
53	53	53
54	54	54
55	55	55
56	56	56
57	57	57
58	58	58
59	59	59
60	60	60
61	61	61
62	62	62
63	63	63
64	64	64
65	65	65
66	66	66
67	67	67
68	68	68
69	69	69
70	70	70
71	71	71
72	72	72
73	73	73
74	74	74
75	75	75
76	76	76
77	77	77
78	78	78
79	79	79
80	80	80
81	81	81
82	82	82
83	83	83
84	84	84
85	85	85
86	86	86
87	87	87
88	88	88
89	89	89
90	90	90
91	91	91
92	92	92
93	93	93
94	94	94
95	95	95
96	96	96
97	97	97
98	98	98
99	99	99
100	100	100

AC ABZ35336;

DT 05-FEB-2003 (first entry)

Human gene expression profile polynucleotide SEQ ID NO 447.

KM Human; artery; endothelium; umbilical; vein; aorta; pulmonary artery;  
KM bronchial epithelium; prostate; muscle; lung fibroblast; osteoblast;  
KM tumour; microarray; genome mapping; antibiotic; antiviral; antifungal;  
KM gene expression; gene; ss.

OS Homo sapiens

PN WO200274979-A2

PD 26-SEP-2002.

PF 20-MAR-2002; 2002WO-US08456.

PR 20-MAR-2001; 2001US-276947P.

PA (ORTH ) ORTHO CLINICAL DIAGNOSTICS INC

PI Wan J, Wang Y;

DR. WPI; 2002-740862/80.

PT New gene expression profile generated from primary, endothelial,

PT pathologies involving alterations of gene expression, e.g. cancer -

PS Example 3; Page 595-496; 850pp; English.

CC The invention relates to a gene expression profile comprising one or more

CC is a coronary artery endothelium, umbilical artery or vein endothelium

endothelium, myoepithrium, microvascular endothelium, keratinocyte  
epithelium, bronchial epithelium, mammary epithelium, prostate  
epithelium, renal cortical epithelium, renal proximal tubule epithelium,  
small airway epithelium, renal epithelium, umbilical artery smooth  
muscle, neonatal dermal fibroblast, pulmonary artery smooth muscle,  
dermal fibroblast, neural progenitor cells, skeletal muscle, astrocytes,  
aortic smooth muscle, mesangial cells, coronary artery smooth muscle,  
bronchial smooth muscle, uterine smooth muscle, lung fibroblast,  
osteoblasts or prostate stromal cell. The gene expression profile is used  
for determining the level of RNA expression for a sample, determining the  
phenotype of a cell and distinguishing cell types. The gene or a protein  
expression profile is useful in identifying disease pathologies  
involving alterations of gene expression. The assessment of expression  
profiles may provide meaningful information with respect to tumour type  
and stage, treatment methods, and prognosis. The gene or protein  
expression profile may also be used for creating microarrays. The  
microarray is useful for genetic and physical mapping of genomes, DNA  
sequencing, genetic or medical diagnosis, genotyping of organisms,  
confirming cell or tissue identifications and in identifying promising  
antibiotics, antiviral or antifungal agents.

Query Match	92.9%	Score 1549.4	DB 24;	length 3382;
Best Local Similarity	97.8%	Pred. No. 0;		
Matches 1626; Conservative	5;	Mismatches	24;	Indels 7; Gaps 6;

QY 1 CACACTGCTCGCTCGGATACTCCAGGCGTCTCCCGTTGCGGCCGCWTCCTWAGAG 60

Db 22 CACACTGCTCGCTTCGGATACCTCCAGGCGTCTCCCGTGGCGGCGCTCCCTGCGCTTAGAG 81

61 GCCAGCCCTGGACACTTGCCTGCCCCCTTCCAGCCCGATTCTGGATCTTCCCTCTGAG 120

D5 82 GCCAGCCTGGACACTTGCCTGCCCCCTTTCACGCCCGAATCTGGGATCCCTCCCTCTGAG 141

121 CCACATCTGGGTCTGCCTTCGACACCACCCCAAGGCTTCCTACCTTGCATGCTTGAG 180

D5 142 CCACATCTGGTCTGCTTCGACACCACCCCAAGCTTCCATCTGCGTCCGAG 201

181 TCTGCCCAAGGGCCCTGTCTCTGGCCATGGCCMAGAAAGGGTCTCTGGGCTGGCA 240

DB 202 TCTGCCCCAGGGGCTCTGTGCTCTGGCCATGGCCAGAGGGGTCCTGGGGCTGGCA 261

241 GCTGGGGCTGTGGCCATTCCTGCTCACTCTGGATTAATCTCGTCTGGGACAGGACCGA 300

Db 262 GCTGGGGGCTGTGGCCATTCTCTCTATCTTTGGATTACTCCGGTCCGGGACAGGAGCGA 321

301 AGGGCAGAGCTTCCTGCGGTGTGGCCCCCAAGTACGCAI CACAGGTGGTACGAGTGC 360

Db 322 AGGGCAGAGCTCCCTGCGGTGTGGCCCCCAGCACGCAI CACAGGTGGCAGCAGTC 381

361 AGTCGCCGTCAGTGGCCCTGGCAGTACCACTAAGAAGGCTCAGTGTGG 420

DB 382 AGTCGCCGTCAGTGGCCCTGGCAGTACGCACTCCTAAGAAAGGCTCCATGAGTGG 741

421 TGGCTCTCTGGAGAGGGTCTGACCTGCTCACAGCTCCCAAGGAGCA 480

DB 442 TGGCTCTCTCTGTCGAGCAGTGGTGCTGTGACCTCTCACCCTCTCCCAAGAGCA 301

481 CCACAAGGAGCTTAIGAGGTCAGCTGGGGCCACCAGCTAGALCTTACTCCGAG 540

D5 502 CCACAGGAGCCTAIGAGGICAGWGGGGCCACCAGCAGTACATCTTACTTCGAGGA 561

• 541 CGCCAGGTCAGCACCTTGAAGGACATCATCCCCACCCAGTACCTCCAGAGGGCTC 600

DB 562 CGCCAGGTCAGCACTTGAAGACATCATCCCCACCACGCTACCTCCAGAGGCTC 621

601 CCAGGGCGACATWGCACCTCCCTCCAACTCAGACAGCCCATCACCTCTCCCGTACATCCG 660

DB 622 CCAGGGGACATTCACATCCCTCAACTCAGCAGACCCTACCTTCCTCCGCTACATCCG 681

661 GCCCACTCTGCCCTCCCTGGACCCAAAGGCTCCACAGCACAGTAC / 20

Db 682 GCCATCTGCTCCCTGACGACCAAGCCCTCTTCCCAACGAGCTCCACTGCACTGTGCAC 741  
QY 721 TGGCTGGGGATCATGTGAGCCCTCAGTGAAGCTCTTGACGCCCAAGCCACTGAGCAACT 780  
Db 742 TGGCTGGGGATCATGTGAGCCCTCAGTGAAGCTCTTGACGCCCAAGCCACTGAGCAACT 801  
QY 781 CGAGGTGCTCTGATCATGTGAGAGAGTGTGTAATCTGCTGTGTAACAATGAGCGCAAGC 840  
Db 802 CGAGGTGCTCTGATCATGTGAGAGAGTGTGTAATCTGCTGTGTAACAATGAGCGCAAGC 860  
QY 841 CTGAGAGCGGCACTTGTGTAAGAGAGCAATGAGTGTGCTGTGTAATGAGAGAGAGGCA 900  
Db 861 CTGAGAGCGGCACTTGTGTAAGAGAGCAATGAGTGTGCTGTGTAATGAGAGAGGCA 920  
QY 901 AGGAGCGCTGACGAGGAGTGAATCTGAGGAGCCCACTCTGCTGCTGAGAGAGTCTGAGT 960  
Db 921 AGGAGCGCTGACGAGGAGTGAATCTGAGGAGCCCACTCTGCTGCTGAGAGAGTCTGAGT 980  
QY 961 ACCTGACGGGCAATTTGAGCTGAGGAGAGATGCTGTGAGGAGCCGCAACAGGCTGTGTGT 1020  
Db 981 ACCTGACGGGCAATTTGAGCTGAGGAGAGATGCTGTGAGGAGCCGCAACAGGCTGTGTGT 1040  
QY 1021 ACACCTGAGCTGAGCTGAGTATGCTCTGAGTCCAAAGCAAGTGAAGAGAACTGAGCTTC 1080  
Db 1041 ACACCTGAGCTGAGCTGAGTATGCTCTGAGTCCAAAGCAAGTGAAGAGAACTGAGCTTC 1100  
QY 1081 GTGTGAGTGGCCCAACCCAGAGAGTCCAGGAGCCGCAACAGCAACTCTGTGAGGAGCACTTG 1140  
Db 1101 GTGTGAGTGGCCCAACCCAGAGAGTCCAGGAGCCGCAACAGCAACTCTGTGAGGAGCACTTG 1160  
QY 1141 CTTTCACTCTGAGCCCAAGCCAGGAGCTGTGAGGAGCCCACTCTTCTGCTCTGAGGAGC 1200  
Db 1161 CTTTCACTCTGAGCCCAAGCCAGGAGCTGTGAGGAGCCCACTCTTCTGCTCTGAGGAGC 1220  
QY 1201 TGGCTGAGGAGCCCTCTCTCCCAATGAGTGAAGAGCACTGAGTGGCCCTCACTTCCAGGA 1260  
Db 1221 TGGCTGAGGAGCCCTCTCTCCCAATGAGTGAAGAGCACTGAGTGGCCCTCACTTCCAGGA 1280  
QY 1261 TGGATCATCATCACTCAAGAGAGAGAGAGCTGCTCTTCCCTGATGAGCTTTGAGAGCCAG 1320  
Db 1281 TGGATCATCATCACTCAAGAGAGAGAGAGCTGCTCTTCCCTGATGAGCTTTGAGAGCCAG 1340  
QY 1321 GCTGACCTGAGGAGCACTCTTCTTCAAGAGACTGTGAGGAGAGGCTGAGGAGCCCACTTGTAT 1380  
Db 1341 GCTGACCTGAGGAGCACTCTTCTTCAAGAGACTGTGAGGAGAGGCTGAGGAGCCCACTTGTAT 1400  
QY 1381 CTTTGAAGCCATCTTCTGTGAGTGTCTTTTGGAGCACTGAGAGTGAAGAGTTTGA 1440  
Db 1401 CTTTGAAGCCATCTTCTGTGAGTGTCTTTTGGAGCACTGAGAGTGAAGAGTTTGA 1460  
QY 1441 CTGCTGTGAGCAATGAGGAGGCTGTGAGGAGCCCTGCAACCAATGAGAGCCGAGCTTGGG 1500  
Db 1461 CTGCTGTGAGCAATGAGGAGGCTGTGAGGAGCCCTGCAACCAATGAGAGCCGAGCTTGGG 1520  
QY 1501 CGAGGCTGTGAGGAGG--TCTGTGAGACTTGTGATGAAATGAGAGCCCTGAGGTTCCACCT 1558  
Db 1521 CGAGGCTGTGAGGAGGCTGTGAGGAGCCCTGATGAAATGAGAGCCCT--GAGTCCACAGCT 1579  
QY 1559 GTTTCCTGAGAGAGTGTCTTCCGAGCCGAGCTTCCCAAGTCTGATGAGAGCACTTTTGTG 1618  
Db 1580 GTTTCCTGAGAGAGTGTCTTCCGAGCCGAGCTTCCCAAGTCT--GATGAGAGAGCA--TCTCTCTG 1636  
QY 1619 CTTTCTGCTGTGTTTGGGTTGGGCACTTTTGGAGAGTTT 1660  
Db 1637 CTTCTCTGCTGTGTTTGGGTTGGGCACTTTTGTGAGAGTT 1678

RESULT 4  
ABST76501  
ID ABST76501 standard; cDNA; 1834 BP.  
XX  
XX ABST76501;  
XX

DT 11-DEC-2002 (first entry)  
XX  
XX cDNA encoding human ovarian cancer marker OV80.  
DE  
XX Human; ovarian cancer; marker; cancer; familial history; brain disorder;  
KW central nervous system disorder; bacterial meningitis; viral meningitis;  
KW Alzheimer's disease; Parkinson's disease; cerebral oedema; hydrocephalus;  
KW brain herniation; inflammation; encephalitis; testicular disorder;  
KW non-tuberculous granulomatous orchitis; connective tissue disorder;  
KW heart disorder; ischemic heart disease; atherosclerosis; neoplasm;  
KW histological type; carcinogenic; ovarian cancer marker; gene; ss.  
XX  
XX Homo sapiens.  
OS  
XX WO200271928-A2.  
PN  
XX 19-SEP-2002.  
PD  
XX 14-MAR-2002; 2002WO-US07826.  
PF  
XX 14-MAR-2001; 2001US-276025P.  
PR 14-MAR-2001; 2001US-276025P.  
PR 10-AUG-2001; 2001US-311732P.  
PR 19-SEP-2001; 2001US-323580P.  
PR 26-SEP-2001; 2001US-324967P.  
PR 26-SEP-2001; 2001US-325102P.  
PR 26-SEP-2001; 2001US-325149P.  
XX  
XX (MILL-) MILLENNium PHARM INC.  
PA  
XX Monahan JE, Gamaavarapu M, Hoersch S, Kamalakar S, Kovalev SG,  
PI Meyers RE, Morrisey MP, Olandt PJ, Sen A, Vieby PO, Mills GB,  
PI Bast RC, Lu K, Schmandt RE, Zhao X, Glatt K,  
XX  
XX MPI, 2002-723277/78.  
DR P-PDB; ABG96402.  
XX  
XX Assessing whether a patient is afflicted with ovarian cancer, useful in  
PT assessing the stage or progression of the disease, comprises comparing  
PT the expression level of a cancer marker in a sample from a patient and  
PT from a non cancer patient -  
XX  
XX Disclosure; Page 384; 481pp; English.  
PS  
XX The present invention relates to a new method for assessing whether a  
XX patient is afflicted with ovarian cancer. The method involves comparing  
CC the expression level of a marker in a patient sample and the normal level  
CC of expression of the marker in a control non-ovarian cancer sample, where  
CC the marker is selected from 363 cancer markers described in the  
CC specification. The method of the invention is useful in diagnosing or  
CC characterizing cancer, in detecting the presence of cancer as early as  
CC possible, and the recurrence of ovarian cancer. The method may also be of  
CC particular use with patients having an enhanced risk of developing  
CC ovarian cancer (e.g. patients having a familial history of ovarian  
CC cancer). The cancer markers may be used in the management and treatment  
CC of e.g. brain and central nervous system disorders (e.g. bacterial and  
CC viral meningitis, Alzheimer's disease or Parkinson's disease), brain  
CC disorders (e.g. cerebral oedema, hydrocephalus or brain herniations),  
CC inflammations (e.g. bacterial or viral meningitis or encephalitis),  
CC testicular disorders (e.g. non-tuberculous granulomatous orchitis),  
CC connective tissue disorders, or heart disorders (e.g. ischemic heart  
CC disease or atherosclerosis). The compositions and methods may also be  
CC used in assessing the histological type of neoplasm associated with  
CC ovarian cancer, monitoring the progression of ovarian cancer,  
CC determining whether ovarian cancer has metastasized or is likely to  
CC metastasize, selecting a composition for inhibiting ovarian cancer,  
CC assessing the ovarian carcinogenic potential of a compound, or  
CC inhibiting ovarian cancer or at risk of developing ovarian cancer. The  
CC present nucleic acid sequence encodes one of the ovarian cancer markers  
CC described in the invention.  
XX  
XX Sequence 1834 BP; 309 A; 621 C; 526 G; 378 T; 0 other;  
SQ

Query Match	92.28	Score 1537.4	DB 24	Length 1833
Best Local Similarity	97.88	Pred. No. 0		
Matches 1625	Conservative 5	Mid. 24	Indels 8	Gaps 7
QY	1	CACACGTGCGCTTCGGATACCTCCAGAGGATCTCCGTTGCGGCGCGTCCCTGACTTAGAG	60	
Db	22	CACACTGCTCGCTTCGGATACCTCCAGAGGATCTCCGTTGCGGCGCGTCCCTGACTTAGAG	81	
QY	61	GCCAGCCTTGAGACCTTGCTGACCCCTTTCCAGCCCGGATTCGAGATCTTCCCTTGAG	120	
Db	82	GCCAGCCTTGAGACCTTGCTGACCCCTTTCCAGCCCGGATTCGAGATCTTCCCTTGAG	141	
QY	121	CCAACTATCTGGGTCCTGCTTCCGACACACCCCAAGGTTCTTACCTTTGGGCTGAG	180	
Db	142	CCAACTATCTGGGTCCTGCTTCCGACACACCCCAAGGTTCTTACCTTTGGGCTGAG	201	
QY	181	TTCTGCCCGAGGGGCGCTTGTCTGGGGCCATAGCCMAAGAGGGGGTCTTGGGGCTGGACA	240	
Db	202	TTCTGCCCGAGGGGCGCTTGTCT-GGCCATAGCCMAAGAGGGGGTCTTGGGGCTGGACA	260	
QY	241	GCTGGGGGCTTGCGCATCTGCTCTATCTTGAATTACTCGGTCGGGACAGAGCGGA	300	
Db	261	GCTGGGGGCTTGCGCATCTGCTCTATCTTGAATTACTCGGTCGGGACAGAGCGGA	320	
QY	301	AGGGGCGAAGCTTCTCGCGGTGGGCCCCCAAGACAGCATCACAGTGGCACACTGC	360	
Db	321	AGGGGCGAAGACCTCCCTCGCGGTGGGCCCCCAAGACAGCATCACAGTGGCACACTGC	380	
QY	361	AGTGGCGCGGTAGTGGCCCTGAGAGGTGAGCATCACTATAGAGCGCTCAATGTGTGG	420	
Db	381	AGTGGCGCGGTAGTGGCCCTGAGAGGTGAGCATCACTATAGAGCGCTCAATGTGTGG	440	
QY	421	TGGCTCTTCGTGTCTGAGCAGTGGGTCTGTCACTGCTCACTTCCCGACGAGCA	480	
Db	441	TGGCTCTTCGTGTCTGAGCAGTGGGTCTGTCACTGCTCACTTCCCGACGAGCA	500	
QY	481	CCACAAGGAACCTATAGAGTCAAGCTGGGGGCCACCAAGTAACTCTTCCGAGGA	540	
Db	501	CCACAAGGAACCTATAGAGTCAAGCTGGGGGCCACCAAGTAACTCTTCTCCGAGGA	560	
QY	541	CGCCAAAGTCAAGACCTCGAAGAGACATCATCCCCCAAGCTACCTCCAGAGAGGCTC	600	
Db	561	CGCCAAAGTCAAGACCTCGAAGAGACATCATCCCCCAAGCTACCTCCAGAGAGGCTC	620	
QY	601	CCAGGGCGACATTGCACTCTTCCAACTGAGAGACCCATCACTTCTTCCGCTACATCCG	660	
Db	621	CCAGGGCGACATTGCACTCTTCCAACTGAGAGACCCATCACTTCTTCCGCTACATCCG	680	
QY	661	GCCCATCTGCTCCTCTGACGACCAAGGCTCTTCCCAAGGCGCTCCACTGACTGTAC	720	
Db	681	GCCCATCTGCTCCTCTGACGACCAAGGCTCTTCCCAAGGCGCTCCACTGACTGTAC	740	
QY	721	TGGCTGGGGGTATGTGGGCCCCCTCGTAGAGGCTCTGAGCGCCAAAGCACTGACAACT	780	
Db	741	TGGCTGGGGGTATGTGGGCCCCCTCGTAGAGGCTCTGAGCGCCAAAGCACTGACAACT	800	
QY	781	CGAGGTGCTCTGATCAGTCTGTGACGCTGTAACTGCTGTAAACAATGACGCGCAAGC	840	
Db	801	CGAGGTGCTCTGATCAGTCTGTGACGCTGTAACTGCTGTAAACAATGACGCGCAAGC	859	
QY	841	CTGAGAGACCGCATTTTGTCCAAAGGACAATGTGTGTGTGCTGTATGTGAGGGGGACA	900	
Db	860	CTGAGAGACCGCATTTTGTCCAAAGGACAATGTGTGTGTGCTGTATGTGAGGGGGACA	919	
QY	901	AGGAGCGCTGCGAAGGGTGATCTGGGGGGCCACTCTCGCGCTGTGAGAGGTCCTGGT	960	
Db	920	AGGAGCGCTGCGAAGGGTGATCTGGGGGGCCACTCTCTCGCTGTGAGAGGTCCTGGT	979	
QY	961	ACCTGACCGGACATTTGTGACTGGAGAGATGACTGTGGGACCGCAACAGGCTGGTGTGT	1020	
Db	980	ACCTGACCGGACATTTGTGACTGGAGAGATGACTGTGGGACCGCAACAGGCTGGTGTGT	1039	
QY	1021	ACACTCTGAGCTTCAGCTATGCTCTTGATCAAAAGCAGGTGACAGAACTCAGGCTTC	1080	

Db	1040	ACACTCTGGGCTCCAGTATAGCTCTCCGTGATCCAAAGCAAGGTGACAGAACTCCAGCTC	1099
Qy	1081	GTCGTGTGCCCCAAACCAGAGAGTCCAGCCCCAGACAACTTCTGTGGCAGCCACTGG	1140
Db	1100	GTCGTGTGCCCCAAACCAGAGAGTCCAGCCCCAGACAACTTCTGTGGCAGCCACTGG	1159
Qy	1141	CCTTCACACTCTGCCCCAGGCCAGGGCTTGTGTAGAGGCCATCTTTTCTGTGCTGTGGGCC	1200
Db	1160	CCTTCACACTCTGCCCCAGGCCAGGGCTTGTGTAGAGGCCATCTTTTCTGTGCTGTGGGCC	1219
Qy	1201	TGCGCTCGGGGCTCTCTCCCATAGGCTCAGCGAGCACTAGAGCTGAGCCCTTAATTCCAGGA	1260
Db	1220	TGCGCTCTGGGGCTCTCTCTCCCATAGGCTCAGCGAGCACTAGAGCTGAGCCCTTAATTCCAGGA	1279
Qy	1261	TGATGTCATCACACTCAAGGACAGAGAGCTGTGTCCTTCCCTGATGAGCTCTTGGACCAAG	1320
Db	1280	TGATGTCATCACACTCAAGGACAGAGAGCTGTGTCCTTCCCTGATGAGCTCTTGGACCAAG	1339
Qy	1321	GCCCTGACTTGAAGCACTCTCTCTTCCCTTCAAGACTCTGTGGGGAGGCTGGGGCCCATCTTGAT	1380
Db	1340	GCCCTGACTTGAAGCACTCTCTCTTCCCTTCAAGACTCTGTGGGGAGGCTGGGGCCCATCTTGAT	1399
Qy	1381	CTTTGAGCCCATCTTCTGTGGGTGTGCTTTTGGGACCATCACTAGAGTCAAGAGTTTGA	1440
Db	1400	CTTTGAGCCCATCTTCTGTGGGTGTGCTTTTGGGACCATCACTAGAGTCAAGAGTTTGA	1459
Qy	1441	CTGCCTGTAGCAATGAGCAGAGCCTCTGGGCCCTCAGCCACATGACACAGGCCATTGGG	1500
Db	1460	CTGCCTGTAGCAATGAGCAGAGCCTCTGGGCCCTCAGCCACATGACACAGGCCATTGGG	1519
Qy	1501	CGAAGTCTCTGGGGAG--TCTCTGGGACCTTGGTATGTAAATGAGGCCCTGGGTGCCACCT	1558
Db	1520	CGAAGTCTCTGGGGAGCTCTGGGACCTTGGTATGTAAATGAGGCCCTGGGTGCCACCT	1578
Qy	1559	GTTTCTGTAGAGACTGCTTCCGCGCCGCGCTTCCAGACTGATGATAGACAACTTTTGTGTC	1618
Db	1579	GTTTCTGTAGAGACTGCTTCCGCGCCGCGCTTCCAGACTGATGATAGACAACTTCTCTGTGC	1635
Qy	1619	CNTTTCCTGTGTTTGTGGGTTGGGCAACTTTTGTGAAGTTT	1660
Db	1636	CCTCTCCCTGTGTTCTGTGGGCTGGGGCCACCTTTGTGACAGCTT	1677
RESULT 5			
ABN95716			
ID	ABN95716	standard; DNA; 1834 BP.	
AC	ABN95716;		
XX			
DT	13-AUG-2002	(first entry)	
XX			
DE	Gene #214	used to diagnose liver cancer.	
XX			
KM	Gene; liver cancer; ds; hepatocellular carcinoma; hepatotropic;		
KW	metastatic liver tumour; cytostatic; expression profile; disease state;		
XX	disease progression; drug toxicity; drug efficacy; drug metabolism.		
OS	Homo sapiens.		
XX			
PN	WO200229103-A2.		
XX			
PD	11-APR-2002.		
XX			
PF	02-OCT-2001; 2001WO-US30589.		
XX			
PR	02-OCT-2000; 2000US-237054P.		
XX			
PA	(GENE-) GENE LOGIC INC.		
XX			
PI	Horne D, Alvares C, Peres-Da-Silva S, Vockley JG;		
XX			
DR	WPI; 2002-426119/45.		

XX diagnosing and detecting the progression of liver cancer,  
 PT hepatocellular carcinoma or metastatic liver tumor in a patient,  
 PT involves detecting the level of expression of two or more genes in a  
 PT liver tissue sample -  
 XX  
 PS claim 1, SEQ ID NO 2214; 298bp; English.  
 XX  
 CC The invention relates to a novel method for diagnosing and detecting the  
 CC progression of liver cancer, hepatocellular carcinoma or metastatic liver  
 CC tumour in a patient, and differentiating metastatic liver cancer from  
 CC hepatocellular carcinoma in a patient, involving detecting the level of  
 CC expression of two or more genes represented in ABN93503-ABN97455 in a  
 CC tissue sample. The method of the invention has hepatotropic, and  
 CC cytostatic activity. The method is useful for diagnosing and detecting  
 CC the progression of liver cancer, hepatocellular carcinoma and metastatic  
 CC liver carcinoma in a patient. The method is useful for identifying  
 CC expression profiles which serve as useful diagnostic markers as well as  
 CC markers that can be used to monitor disease states, disease progression,  
 CC drug toxicity, drug efficacy and drug metabolism.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIP0  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 CC  
 SQ Sequence 1834 BP; 309 A; 621 C; 526 G; 378 T; 0 other;  
 Query Match 92.2%; Score 1537.4; DB 24; Length 1834;  
 Best Local Similarity 97.8%; Pred. No. 0;  
 Matches 1655; Conservative 5; Mismatches 24; Indels 8; Gaps 7;

QY 661 GCCCATCTGCTCCCTGACGACCAAGCCTCTTCCCAAGGCTCCACTGCACTGTAC 720  
 DB 681 GCCCATCTGCTCCCTGACGACCAAGCCTCTTCCCAAGGCTCCACTGCACTGTAC 740  
 QY 721 TGGCTGGGGTCAATGAGGCCCCCTCAGTAGCCTCTGACGCCCAAGCCATCAGCACT 780  
 DB 741 TGGCTGGGGTCAATGAGGCCCCCTCAGTAGCCTCTGACGCCCAAGCCATCAGCACT 800  
 QY 781 CGAGGCTCTGATCAGTCGAGACCGGTAATCTGCTGTATCAACATCGACCCCAAC 840  
 DB 801 CGAGGCTCTGATCAGTCGAGACCGGTAATCTGCTGTATCAACATCGACCCCAAC 859  
 QY 841 CTGAGAGGCGGCACTTTGTCGAAGAGACATGATGATGCTGATGAGGAGGAGCA 900  
 DB 860 CTGAGAGGCGGCACTTTGTCGAAGAGACATGATGATGCTGATGAGGAGGAGCA 919  
 QY 901 AGGAGCGCTGCAAGGGTGAATCTGAGGGGCGCACTCTGCTGCTGTGAGGGTCTGTGT 960  
 DB 920 AGGAGCGCTGCAAGGGTGAATCTGAGGGGCGCACTCTGCTGCTGTGAGGGTCTGTGT 979  
 QY 961 ACCTGACGGGCAATGATGAGCTGGGAGATGCTGTGAGGGGCGCAAGCCTGTGTGT 1020  
 DB 980 ACCTGACGGGCAATGATGAGCTGGGAGATGCTGTGAGGGGCGCAAGCCTGTGTGT 1039  
 QY 1021 ACACTCTGGGCTCCAGCTATGCTCTGGAATCAAGACAGGTGACAGAACTCCAGGCTC 1080  
 DB 1040 ACACTCTGGGCTCCAGCTATGCTCTGGAATCAAGACAGGTGACAGAACTCCAGGCTC 1099  
 QY 1081 GTGTGGTGGCCCAACCCAGAGTCCAGGCCGAGACAGCAACTCTGTGTGACGCACTTG 1140  
 DB 1100 GTGTGGTGGCCCAACCCAGAGTCCAGGCCGAGACAGCAACTCTGTGTGACGCACTTG 1159  
 QY 1141 CTTTGAAGCTGAGCCCAAGCCGAGGCTGTGAGGCCATCTTTTCTGCTCTGAGGCC 1200  
 DB 1160 CTTTGAAGCTGAGCCCAAGCCGAGGCTGTGAGGCCATCTTTTCTGCTCTGAGGCC 1219  
 QY 1201 TGGCTCTGGGCTCTCTCTCCCATGAGTCAAGGACACTGAGAGCTGAGCCTTCCAGGA 1260  
 DB 1220 TGGCTCTGGGCTCTCTCTCCCATGAGTCAAGGACACTGAGAGCTGAGCCTTCCAGGA 1279  
 QY 1261 TGGATGATCACTCAAGAGACAGGAGCTGTCTCTCTGATGAGCTTTTGAACCCAG 1320  
 DB 1280 TGGATGATCACTCAAGAGACAGGAGCTGTCTCTCTGATGAGCTTTTGAACCCAG 1339  
 QY 1321 GCTTGAATGAGCACTCTCTCTTCAAGATCTTGGGAGGCTTGGGCCCCCATCTTGAT 1380  
 DB 1340 GCTTGAATGAGCACTCTCTCTTCAAGATCTTGGGAGGCTTGGGCCCCCATCTTGAT 1399  
 QY 1381 CTTTGAAGCCATCTTCTGGGATGCTTTTGGGACATCACTGAGAGTCAGAGGTTTA 1440  
 DB 1400 CTTTGAAGCCATCTTCTGGGATGCTTTTGGGACATCACTGAGAGTCAGAGGTTTA 1459  
 QY 1441 CTGCTGTAGCAATGAGCCAGAGCCTCTGAGCCCTTCAACCAAGGAGCCAGCCATTTGGS 1500  
 DB 1460 CTGCTGTAGCAATGAGCCAGAGCCTCTGAGCCCTTCAACCAAGGAGCCAGCCATTTGGS 1519  
 QY 1501 CGAGTCTCTGGGAGG--TCTGTGAGACCTTGTGTATGAAAATGAGCCCTGTGCTCCACT 1558  
 DB 1520 CGAGTCTCTGGGAGCCTCTGAGACCTTGTGTATGAAAATGAGCCCTGTGCTCCACT 1578  
 QY 1559 GTTTCCTGGAAGCTCTTCCGAGCCGAGCTTCCAGACCTGATGATGAGCACTTTTGGC 1618  
 DB 1579 GTTTCCTGGAAGCTCTTCCGAGCCGAGCTTCCAGACCTGATGATGAGCACTTCTCTGC 1635  
 QY 1619 CTTTCTCTGTGTTTGTGAGGCACTTTTGGAAATTT 1660  
 DB 1636 CTTCTCCCTGTGTCTGTGGGCTGGGCGCACTTTGTGAGCTT 1677

RESULT 6  
 ABK12241  
 ID ABK12241 standard; cDNA; 1834 BP.

```

XX ABK12241;
AC 18-JUN-2002 (first entry)
XX
XX cDNA encoding human prostatic protein.
XX
XX Prostatic; human; malignant; cancer; ovarian cancer; sr; gene;
XX breast cancer; prostate cancer; lung cancer; colon cancer.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
XX CDS 229..1260
XX FT /*tag= a
XX FT /product= "Prostatic protein"
XX
XX MO200221133-A2.
XX
XX 14-MAR-2002.
XX
XX 07-SEP-2001; 2001MO-US27718.
XX
XX 07-SEP-2000; 2000US-231166P.
XX
XX (BGHM ) BRIGHAM & WOMENS HOSPITAL INC.
XX
XX Mok SC, Wong K;
XX
XX MPI; 2002-292285/33.
XX P-PSDB; AAU78547.
XX
XX Diagnosing cancer or susceptibility to it, useful particularly for
XX ovarian cancer, comprises detecting increased levels of prostatic in
XX blood or tissue
XX
XX Disclosure; Page 32-34; 36pp; English.
XX
XX This invention relates to a novel method for determining if a human has,
XX or is likely to develop, a malignant growth. The method comprises
XX measuring the concentration of prostatic protein or mRNA, in a sample
XX and comparing the result with one or more controls, where a level of
XX prostatic expression that is significantly higher than in controls is
XX indicative of cancer, or increased susceptibility. The assay is
XX specifically used to detect ovarian cancer or a predisposition to it,
XX but may also be applicable to breast, prostatic, lung and colonic
XX cancers. Using the method of the invention it was shown that in ovarian
XX cancer, the highest levels of prostatic were found in stage II disease,
XX suggesting that this marker is suitable for early detection. The
XX present sequence represents the cDNA encoding the human prostatic
XX protein used in the method of the invention as a marker for early
XX detection of cancer or susceptibility to cancer.
XX
XX Sequence 1834 BP; 309 A; 621 C; 526 G; 378 T; 0 other;
XX
XX Query Match 92.2%; Score 1537.4; DB 24; Length 1834;
XX Best Local Similarity 97.8%; Pred. No. 0;
XX Matches 162; Conservative 5; Mismatches 24; Indels 8; Gaps 7;
XX
QY 1 CACACTGCTGCTTGGATACCTCCAGGCTCTCCGTTGCGGCGCTCCCTGCTTGAAG 60
DB 22 CACACTGCTGCTTGGATACCTCCAGGCTCTCCGTTGCGGCGCTCCCTGCTTGAAG 81
QY 61 GCCAGCTTGGACACTTGTCTGCTCCCTTCCAGCCCGGATTTGGGATCTTCTCTGAG 120
DB 82 GCCAGCTTGGACACTTGTCTGCTCCCTTCCAGCCCGGATTTGGGATCTTCTCTGAG 141
QY 121 CCAACATCTGGGCTCTGCTTGCACACACCCCAAGGCTTCTTCTGCTTGGCTGAG 180
DB 142 CCAACATCTGGGCTCTGCTTGCACACACCCCAAGGCTTCTTCTGCTTGGCTGAG 201
QY 181 TCTGCCCAAGGGGCTTGTCTCTGAGGCTATGCGMAGAGGGGCTCTGAGGCTTGGACA 240

```

```

DB 202 TCTGCCCAAGGGGCTTGTCTCTGAGGCTATGCGMAGAGGGGCTCTGAGGCTTGGACA 260
QY 241 GCTGGGGGCTGTGGCCATCTGCTCTATCTTGGATTACTCCGGTGGGGACAGAGCGGA 300
DB 261 GCTGGGGGCTGTGGCCATCTGCTCTATCTTGGATTACTCCGGTGGGGACAGAGCGGA 320
QY 301 AGGGGCAAGAGCTTCTGGGGTGTGGCCCCCAAGCAACATCAAGTGGACAGAGTGC 360
DB 321 AGGGGCAAGAGCTTCTGGGGTGTGGCCCCCAAGCAACATCAAGTGGACAGAGTGC 380
QY 361 AGTGGCCGCTCAATGGGCTTGGCAAGTCAAGTCACTTGAAGGCTCAATGTGTGG 420
DB 381 AGTGGCCGCTCAATGGGCTTGGCAAGTCAAGTCACTTGAAGGCTCAATGTGTGG 440
QY 421 TGGCTCTCTGCTGCTGACAGAGAGGAGTGTGCTGAGTGTCACTGTTCCGACAGGA 480
DB 441 TGGCTCTCTGCTGCTGACAGAGAGGAGTGTGCTGAGTGTCACTGTTCCGACAGGA 500
QY 481 CCAGAAAGAGCTATGAGGTCAAGCTGGGGGCCCAAGCTAGACTCTCTACTCCAGGA 540
DB 501 CCAGAAAGAGCTATGAGGTCAAGCTGGGGGCCCAAGCTAGACTCTCTACTCCAGGA 560
QY 541 CGCCAAAGTCAAGACCTTGAAGACATCATCCGCCCAAGCTTCACTTCAAGAGGGCTC 600
DB 561 CGCCAAAGTCAAGACCTTGAAGACATCATCCGCCCAAGCTTCACTTCAAGAGGGCTC 620
QY 601 CCAGGGCGCAATTTGCACTCTCTCAACTCAGAGAACCATCACTTCTCCGCTACATCCG 660
DB 621 CCAGGGCGCAATTTGCACTCTCTCAACTCAGAGAACCATCACTTCTCCGCTACATCCG 680
QY 661 GCCCATCTGCTCTCCCTGACAGCAAGCGCTCTTCCCAAGCGGCTCACTGCACTGTAC 720
DB 681 GCCCATCTGCTCTCCCTGACAGCAAGCGCTCTTCCCAAGCGGCTCACTGCACTGTAC 740
QY 721 TGGCTGGGCTCATGTGGCCCCCTCACTGAGCTCTTCAAGCCCAAGCCACTGAGAACT 780
DB 741 TGGCTGGGCTCATGTGGCCCCCTCACTGAGCTCTTCAAGCCCAAGCCACTGAGAACT 800
QY 781 CGAGTGGCTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 840
DB 801 CGAGTGGCTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 860
QY 841 CTGAGAGCGCGCACTTGTTCAGAGAGACATGATGTGTGTGTGTGTGTGTGTGTGTGT 900
DB 860 CTGAGAGCGCGCACTTGTTCAGAGAGACATGATGTGTGTGTGTGTGTGTGTGTGTGT 919
QY 901 AGAGCGCTGCTCCAGGGTACTCTGGGGGCCCACTCTCTGCTGTGGAGGGCTCTGTGT 960
DB 920 AGAGCGCTGCTCCAGGGTACTCTGGGGGCCCACTCTCTGCTGTGGAGGGCTCTGTGT 979
QY 961 ACCTGACGGGCAATTGAGAGTGGAGAGATGCTGTGGGGGCCCAAGGCTGTGTGT 1020
DB 980 ACCTGACGGGCAATTGAGAGTGGAGAGATGCTGTGGGGGCCCAAGGCTGTGTGT 1039
QY 1021 ACACTGTGGCTTCACTATGCTCTCTGATTCAGAAAGAGTGAAGAACTTCCAGCTTC 1080
DB 1040 ACACTGTGGCTTCACTATGCTCTCTGATTCAGAAAGAGTGAAGAACTTCCAGCTTC 1099
QY 1081 GTGTGTGCTCCCAAAACCCAGAGATCCAGGCCGACAGCAACTCTGTGGAGAGCACTGG 1140
DB 1100 GTGTGTGCTCCCAAAACCCAGAGATCCAGGCCGACAGCAACTCTGTGGAGAGCACTGG 1159
QY 1141 CTTTCAAGCTTGGCCGAGCCAGGGCTGTGGAGGCCCACTTTTCTGCTGCTGAGGCC 1200
DB 1160 CTTTCAAGCTTGGCCGAGCCAGGGCTGTGGAGGCCCACTTTTCTGCTGCTGAGGCC 1219
QY 1201 TGGCTTGGGCTCTCTCTTCCCAATGCTCAAGCAGACATGAGCTGAGCTTCTTCAAGA 1260
DB 1220 TGGCTTGGGCTCTCTCTTCCCAATGCTCAAGCAGACATGAGCTGAGCTTCTTCAAGA 1279
QY 1261 TGAATGATCACTCAAGAGACAGAGCTGTGCTTCTCTGATGTGCTTGTGAACCAAG 1320
DB 1280 TGAATGATCACTCAAGAGACAGAGCTGTGCTTCTCTGATGTGCTTGTGAACCAAG 1339

```



```

QY 1321 GCCTGACCTGAGCCACTCTCTCTTGAAGACTCTGCGGAGGCTGGGGCCCATCTTGAT 1380
Db 1340 GCGCTGACTGAGCCACTCTCTCTTGAAGACTCTGCGGAGGCTGGGGCCCATCTTGAT 1399
QY 1381 CTTTGAGCCCATCTCTGCGGTGCTCTTTTGGGACCATACAGAGTGAAGATTTA 1440
Db 1400 CTTTGAGCCCATCTCTGCGGTGCTCTTTTGGGACCATACAGAGTGAAGATTTA 1459
QY 1441 CTGCTGTGACATGAGCCAGAGCCCTGAGCCCTCAGCCATGAGCCAGCCATTTGAG 1500
Db 1460 CTGCTGTGACATGAGCCAGAGCCCTGAGCCCTCAGCCATGAGCCAGCCATTTGAG 1519
QY 1501 CGAGNTCTGGGGAG--TCCTGGACCTTGGTATGAAATGAGCCCTGGGTCCACCT 1558
Db 1520 CGAGCTCTGGGGAGCTCTGGGACCTTGGCTATGAAATGAGCCCT--GGCTCCACCT 1578
QY 1559 GTTTCTGGAAGATGCTCTCCCGGCGGCTCCCAAGCTGATGAGACACTTTTGTG 1618
Db 1579 GTTTCTGGAAGATGCTCTCCCGGCGGCTCCCAAGCT--GATGAGACACT--TCTCTCTGC 1635
QY 1619 CTTTCTCCCTGTGTTTGGGTGGGCAACTTTTGAAGTTT 1660
Db 1636 CCTCTCCCTGTGTTTGGGTGGGCGCCACCTTTGTGAGCTT 1677

```

RESULT 7  
ABL67949  
ID ABL67949 standard; DNA: 1834 BP.

XX ABL67949;

DT 15-MAY-2002 (first entry)

XX Ovary cancer related gene sequence SEQ ID NO:6286.

XX Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid;  
XX stomach; lung; prostate; pancreas; carcinoma; antitumor; cancerous;  
XX cytostatic; gene therapy; antineoplastic; Wilm's tumour; adenocarcinoma;  
XX gene; de.

OS Homo sapiens.

PN WO200194629-A2.

PD 13-DEC-2001.

XX 30-MAY-2001; 2001WO-US10838.

XX 05-JUN-2000; 2000US-209473P.

XX 18-SEP-2000; 2000US-209531P.

XX 20-SEP-2000; 2000US-233617P.

XX 20-SEP-2000; 2000US-234009P.

XX 20-SEP-2000; 2000US-234052P.

XX 22-SEP-2000; 2000US-234509P.

XX 22-SEP-2000; 2000US-234567P.

XX 25-SEP-2000; 2000US-234923P.

XX 25-SEP-2000; 2000US-234924P.

XX 25-SEP-2000; 2000US-235077P.

XX 25-SEP-2000; 2000US-235134P.

XX 25-SEP-2000; 2000US-235280P.

XX 26-SEP-2000; 2000US-235637P.

XX 27-SEP-2000; 2000US-235638P.

XX 27-SEP-2000; 2000US-235711P.

XX 27-SEP-2000; 2000US-235720P.

XX 27-SEP-2000; 2000US-235840P.

XX 28-SEP-2000; 2000US-235863P.

XX 28-SEP-2000; 2000US-236028P.

XX 28-SEP-2000; 2000US-236032P.

XX 28-SEP-2000; 2000US-236033P.

```

PR 28-SEP-2000; 2000US-236034P.
PR 28-SEP-2000; 2000US-236109P.
PR 28-SEP-2000; 2000US-236111P.
PR 29-SEP-2000; 2000US-236842P.
PR 29-SEP-2000; 2000US-236891P.
PR 02-OCT-2000; 2000US-237172P.
PR 02-OCT-2000; 2000US-237173P.
PR 02-OCT-2000; 2000US-237278P.
PR 02-OCT-2000; 2000US-237294P.
PR 02-OCT-2000; 2000US-237295P.
PR 02-OCT-2000; 2000US-237316P.
PR 03-OCT-2000; 2000US-237425P.
PR 03-OCT-2000; 2000US-237598P.
PR 03-OCT-2000; 2000US-237604P.
PR 03-OCT-2000; 2000US-237606P.
PR 03-OCT-2000; 2000US-237608P.
PR 01-NOV-2000; 2000US-244867P.
PR 01-NOV-2000; 2000US-245084P.

```

XX (AVAL-) AVALON PHARM.

PI Young PE, Augustus M, Carter KC, Ebner R, Endress G, Horrigan S;

XX Soppet DR, Weaver Z;

XX MPI; 2002-188264/24.

PT Screening for anti-neoplastic agent involves exposing cells to a

PT chemical agent to be tested for anti-neoplastic activity, and

PT determining a change in expression of a gene of a signature gene set -

PS Claim 1; SEQ ID 6286; 44pp; English.

XX The present invention describes a method (M1) for screening for an

XX anti-neoplastic agent. The method involves exposing cells to a chemical

XX agent to be tested for anti-neoplastic activity, determining a change in

XX expression of at least one gene (I) of a signature gene set, where (I)

XX comprises a sequence (S) selected from 8447 sequences (given in ABL6164

XX to ABL70110), or is at least 95% identical to (S), where a change in

XX expression is indicative of anti-neoplastic activity. (I) has cytostatic

XX activity and can be used in gene therapy. M1 can be used for screening

XX an anti-neoplastic agent, and can be used for producing a product which

XX is the data collected with respect to the anti-neoplastic agent as a

XX result of M1, and the data is sufficient to convey the chemical

XX structure and/or properties of the agent. M1 can be used in the

XX treatment of cancer such as colon, breast, stomach, lung, thyroid,

XX oesophageal, ovarian, kidney, prostate or pancreatic cancer,

XX adenocarcinoma, carcinoma, clear cell cancer, infiltrating ductal cancer,

XX infiltrating lobular cancer, squamous cell carcinoma, neuroendocrine

XX carcinoma, papillary carcinoma and Wilm's tumour.

XX Sequence 1834 BP; 309 A; 621 C; 526 G; 378 T; 0 other;

Query Match 92.2%; Score 1537.4; DB 24; Length 1834;

Best Local Similarity 97.8%; Pred. No. 0;

Matches 1625; Conservative 5; Mismatches 24; Indels 8; Gaps 7;

```

QY 1 CACACTGCTCGCTTGGATCTCCAGGCGTCCCGGCGGCGGCTCCCTGCTTAAG 60
Db 22 CACACTGCTCGCTTGGATCTCCAGGCGTCCCGGCGGCGGCTCCCTGCTTAAG 81
QY 61 GCCAGCCTTGGACACTTTCGCTTCCAGCCCGGATTCGAGATCTTCCCTGAG 120
Db 82 GCCAGCCTTGGACACTTTCGCTTCCAGCCCGGATTCGAGATCTTCCCTGAG 141
QY 121 CCAACATCTGGGTCTGCTTGGACACACCCCAAGGCTTCTTACCTTGGCTGAG 180
Db 142 CCAACATCTGGGTCTGCTTGGACACACCCCAAGGCTTCTTACCTTGGCTGAG 201
QY 181 TTGAGCCAGAGGCGCTTGTCTGGGCGATGGCGAAGAGGGGCTCGGGGCTGGGCA 240
Db 202 TTGAGCCAGAGGCGCTTGTCT--GGCATGGCCCAAGAGGGGCTCGGGGCTGGGCA 260
QY 241 GCTGGGGGCTGTGGCAATTCGTCTATCTTGATTTACTCCGGTGGGAGACAGAGCGGA 300

```

```

Db      ||| 261 GCTGGGGGCTGTGGCATTCGTCTATCTTGGATTACTCCGGTCGGGGCAGAGAGGGA 320
Qy      ||| 301 AGGGGCAAGAGCTTCCTGGGCTGTGGCCCCCAAGACGCGATCAGAGTGGGAGAGTGC 360
Db      ||| 321 AGGGGCAAGAGCTTCCTGGGCTGTGGCCCCCAAGACGCGATCAGAGTGGGAGAGTGC 380
Qy      ||| 361 AGTGGCGGCTCAGTGGGCTGTGGAGGTGAGCATCATCCTATGAAGGCTCCATGTGTGG 420
Db      ||| 381 AGTGGCGGCTCAGTGGGCTGTGGAGGTGAGCATCATCCTATGAAGGCTCCATGTGTGG 440
Qy      ||| 421 TGGCTCTCTGTGTCTGAGCAGTGGGTGTGTGTGAGCTGTGACTGTCTTCCCAAGGAGA 480
Db      ||| 441 TGGCTCTCTGTGTCTGAGCAGTGGGTGTGTGTGAGCTGTGACTGTCTTCCCAAGGAGA 500
Qy      ||| 481 CCACAAAGAAAGCTTATGAGGTCAAGCTGGGGGGCCACAGCTAGACTCTCCATCTCCAGGA 540
Db      ||| 501 CCACAAAGAAAGCTTATGAGGTCAAGCTGGGGGGCCACAGCTAGACTCTCCATCTCCAGGA 560
Qy      ||| 541 CGCGAAGGTGAGCACCCTGAAGGACATCATCCCCCAGCTACCTCCAGAGAGGCTC 600
Db      ||| 561 CGCGAAGGTGAGCACCCTGAAGGACATCATCCCCCAGCTACCTCCAGAGAGGCTC 620
Qy      ||| 601 CGAGGGGAGCATTGACCTCTCCAACTCAGCAGACCCCATCATCTTCTCCGCTACATCG 660
Db      ||| 621 CGAGGGGAGCATTGACCTCTCCAACTCAGCAGACCCCATCATCTTCTCCGCTACATCG 680
Qy      ||| 661 GCCCATCTGCTCCCTGAGCAGCCAAAGGCTCTTCCCAAGGGCTCCAGCTGCACTGTAC 720
Db      ||| 681 GCCCATCTGCTCCCTGAGCAGCCAAAGGCTCTTCCCAAGGGCTCCAGCTGCACTGTAC 740
Qy      ||| 721 TGGCTGGGGTCAATGTGGCCCCCTCAGTGAAGCTCTGACGCCCAAGCCATGAGCAACT 780
Db      ||| 741 TGGCTGGGGTCAATGTGGCCCCCTCAGTGAAGCTCTGACGCCCAAGCCATGAGCAACT 800
Qy      ||| 781 CGAGTGGCTCTGATCAGTGTGTGAGAGCTGTGTAACTGCTGTACATCAGTGCAGC 840
Db      ||| 801 CGAGTGGCTCTGATCAGTGTGTGAGAGCTGTGTAACTGCTGTACATCAGTGCAGC 860
Qy      ||| 841 CTGAGAGAGCCGCACTTGTCTCAAGAGACATGCTGTGTGCTGTGTGTGTGTGTGTGTGT 900
Db      ||| 860 CTGAGAGAGCCGCACTTGTCTCAAGAGACATGCTGTGTGCTGTGTGTGTGTGTGTGTGT 919
Qy      ||| 901 AGGAGCGCTGACAGGAGTGACTGTGGGGGCCCATCTCTCGGCCCTGTGAGAGGTCTGTGT 960
Db      ||| 920 AGGAGCGCTGACAGGAGTGACTGTGGGGGCCCATCTCTCGGCCCTGTGAGAGGTCTGTGT 979
Qy      ||| 961 ACCTGACGGGCAATTGTGAGCTGGGGAGATGCTGTGTGGGGCCCGCAGACAGGCTGTGTGT 1020
Db      ||| 980 ACCTGACGGGCAATTGTGAGCTGGGGAGATGCTGTGTGGGGCCCGCAGACAGGCTGTGTGT 1039
Qy      ||| 1021 ACACCTGGGCTCCAGCTATGAGTCCCTCGATCCCAACAGAGTGAAGAACTCCAGGCTC 1080
Db      ||| 1040 ACACCTGGGCTCCAGCTATGAGTCCCTCGATCCCAACAGAGTGAAGAACTCCAGGCTC 1099
Qy      ||| 1081 GTGTGTGGCCCAAAACCGAGAGTCCAGCCCGACAGCAACTCTGTGGAGCCACTGTGG 1140
Db      ||| 1100 GTGTGTGGCCCAAAACCGAGAGTCCAGCCCGACAGCAACTCTGTGGAGCCACTGTGG 1159
Qy      ||| 1141 CTTTCACTGTGCCCCAGGCCGCTTGTGAGGCCCATCTTTTCTGTGCTGTGGGCC 1200
Db      ||| 1160 CTTTCACTGTGCCCCAGGCCGCTTGTGAGGCCCATCTTTTCTGTGCTGTGGGCC 1219
Qy      ||| 1201 TGGCTCTGGGCTCTCTCTCCCATGAGCTCAGCCAGCACTAGAGCTGGGCCCTTACAGGA 1260
Db      ||| 1220 TGGCTCTGGGCTCTCTCTCCCATGAGCTCAGCCAGCACTAGAGCTGGGCCCTTACAGGA 1279
Qy      ||| 1261 TGGATGATCACTCAAGAGCAGAGAGCTGTCTTCTCCGATGAGCTTGTGAGCCAGG 1320
Db      ||| 1280 TGGATGATCACTCAAGAGCAGAGAGCTGTCTTCTCCGATGAGCTTGTGAGCCAGG 1339
Qy      ||| 1321 GCCTGACTTGAAGCACTTCCTTCTTCAAGACTTGTGGGAGGCTGGGGCCCATCTTGTAT 1380

```

```

Db      ||| 1340 GCCTGACTTGAGCCACTCTCTCTTACAGACTCTGTGGGAGGCTGGGCCCATCTTGTAT 1399
Qy      ||| 1381 CTTTGAAGCCATTTCTTGGGTGTGCTTTTGGACATCATCTGAGAGTCAAGAGTTTAA 1440
Db      ||| 1400 CTTTGAAGCCATTTCTTGGGTGTGCTTTTGGACATCATCTGAGAGTCAAGAGTTTAA 1459
Qy      ||| 1441 CTGCTGTGAGCAATGAGCCAGAGCTCTGTGCCCCCTCAGCTGATGAGCAGTGTGGS 1500
Db      ||| 1460 CTGCTGTGAGCAATGAGCCAGAGCTCTGTGCCCCCTCAGCTGATGAGCAGTGTGGS 1519
Qy      ||| 1501 CGAGTCTGGGAGGCTCTGTGAGGCTGTGTGTAAGAAATAGAGCCCTGGGTTCCAGCT 1558
Db      ||| 1520 CGAGTCTGGGAGGCTCTGTGAGGCTGTGTGTAAGAAATAGAGCCCTGGGTTCCAGCT 1578
Qy      ||| 1559 GTTTCCTGAGAGCTGCTTCCGCGCCGCTTCCAGCTGATGAGCAGATTTTGTGC 1618
Db      ||| 1579 GTTTCCTGAGAGCTGCTTCCGCGCCGCTTCCAGCTGATGAGCAGACT-GATGAGCAGA-TCCTCTGC 1635
Qy      ||| 1619 CMTTTCCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1660
Db      ||| 1636 CMTTTCCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1677

RESULT 8
ABL68248
ID      ||| ABL68248 standard; DNA; 1834 BP.
XX
AC      ||| ABL68248;
XX
DT      ||| 15-MAY-2002 (first entry)
XX
DE      ||| Kidney cancer related gene sequence SEQ ID NO:6585.
XX
KW      ||| Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid;
KW      ||| stomach; lung; prostate; pancreas; carcinoma; antitumour; Cancerous;
KW      ||| cytotoxic; gene therapy; antineoplastic; Wilms tumour; adenocarcinoma;
KW      ||| gene; ds.
XX
OS      ||| Homo sapiens.
XX
PN      ||| WO200194629-A2.
XX
PD      ||| 13-DEC-2001.
XX
PF      ||| 30-MAY-2001; 2001WO-US10838.
XX
PR      ||| 05-JUN-2000; 2000US-209473P.
PR      ||| 05-JUN-2000; 2000US-209511P.
PR      ||| 18-SEP-2000; 2000US-23133P.
PR      ||| 18-SEP-2000; 2000US-233617P.
PR      ||| 20-SEP-2000; 2000US-234009P.
PR      ||| 20-SEP-2000; 2000US-234034P.
PR      ||| 20-SEP-2000; 2000US-234052P.
PR      ||| 22-SEP-2000; 2000US-234509P.
PR      ||| 22-SEP-2000; 2000US-234567P.
PR      ||| 25-SEP-2000; 2000US-234923P.
PR      ||| 25-SEP-2000; 2000US-234924P.
PR      ||| 25-SEP-2000; 2000US-235072P.
PR      ||| 25-SEP-2000; 2000US-235082P.
PR      ||| 25-SEP-2000; 2000US-235134P.
PR      ||| 25-SEP-2000; 2000US-235280P.
PR      ||| 26-SEP-2000; 2000US-235637P.
PR      ||| 26-SEP-2000; 2000US-235638P.
PR      ||| 27-SEP-2000; 2000US-235711P.
PR      ||| 27-SEP-2000; 2000US-235720P.
PR      ||| 27-SEP-2000; 2000US-235840P.
PR      ||| 27-SEP-2000; 2000US-235840P.
PR      ||| 28-SEP-2000; 2000US-236028P.
PR      ||| 28-SEP-2000; 2000US-236032P.
PR      ||| 28-SEP-2000; 2000US-236033P.
PR      ||| 28-SEP-2000; 2000US-236034P.
PR      ||| 28-SEP-2000; 2000US-236109P.
PR      ||| 28-SEP-2000; 2000US-236111P.

```



PR 29-SEP-2000; 2000US-236842P.  
 PR 29-SEP-2000; 2000US-236891P.  
 PR 02-OCT-2000; 2000US-237172P.  
 PR 02-OCT-2000; 2000US-237173P.  
 PR 02-OCT-2000; 2000US-237278P.  
 PR 02-OCT-2000; 2000US-237294P.  
 PR 02-OCT-2000; 2000US-237295P.  
 PR 02-OCT-2000; 2000US-237316P.  
 PR 03-OCT-2000; 2000US-237425P.  
 PR 03-OCT-2000; 2000US-237598P.  
 PR 03-OCT-2000; 2000US-237604P.  
 PR 03-OCT-2000; 2000US-237606P.  
 PR 03-OCT-2000; 2000US-237608P.  
 PR 01-NOV-2000; 2000US-244867P.  
 PR 01-NOV-2000; 2000US-245084P.

## (AVAL-) AVALON PHARM.

Young PE, Augustus M, Carter KC, Ebner R, Endress G, Horrigan S,  
 Soppet DR, Weaver Z;

WPI; 2002-188264/24.

Screening for anti-neoplastic agent involves exposing cells to a  
 chemical agent to be tested for anti-neoplastic activity, and  
 determining a change in expression of a gene of a signature gene set -

Claim 1, SEQ ID 6585; 44pp; English.

The present invention describes a method (M1) for screening for an  
 anti-neoplastic agent. The method involves exposing cells to a chemical  
 agent to be tested for anti-neoplastic activity, determining a change in  
 expression of at least one gene (I) of a signature gene set, where (II)  
 comprises a sequence (S) selected from 8447 sequences (given in AB161664  
 to AB170110), or is at least 95% identical to (S), where a change in  
 expression is indicative of anti-neoplastic activity. (I) has cytoskeletal  
 activity and can be used in gene therapy. M1 can be used for screening  
 an anti-neoplastic agent, and can be used for producing a product which  
 is the data collected with respect to the anti-neoplastic agent as a  
 result of M1, and the data is sufficient to convey the chemical  
 structure and/or properties of the agent. M1 can be used in the  
 treatment of cancer such as colon, breast, stomach, lung, thyroid,  
 oesophageal, ovarian, kidney, prostate or pancreatic cancer,  
 adenocarcinoma, carcinoma, clear cell cancer, infiltrating ductal cancer,  
 CC infiltrating lobular cancer, squamous cell carcinoma, neuroendocrine  
 carcinoma, papillary carcinoma and Wilms' tumour.

Sequence 1834 BP; 309 A; 621 C; 526 G; 378 T; 0 other;

Query Match 92.2%; Score 1537.4; DB 24; Length 1834;

Best Local Similarity 97.8%; Pred. No. 0;

Matches 1655; Conservative 5; Mismatches 24; Indels 8; Gaps 7;

QY 1 CACACTGCTGCTTGGATGATCCAGGCGTCTCCGCTGCGCGCTCCCTGCTTAAAG 60  
 DB 22 CACACTGCTGCTTGGATGATCCAGGCGTCTCCGCTGCGCGCTCCCTGCTTAAAG 81  
 QY 61 GGCAGCTTGGACATCTGCTGCGCTTTCAGCCCGGATTCGGATCTTCCCTCTAG 120  
 DB 82 GGCAGCTTGGACATCTGCTGCGCTTTCAGCCCGGATTCGGATCTTCCCTCTAG 141  
 QY 121 CCACATCTGCTGCTTTCAGACACACACCCCAAGGCTTCTCACTTCCGCTGAG 180  
 DB 142 CCACATCTGCTGCTTTCAGACACACCCCAAGGCTTCTCACTTCCGCTGAG 201  
 QY 181 TGTGCTCCCAAGGCGCTTCTCTGCGCATGCGCMAAGAGGAGGCTTGGAGCA 240  
 DB 202 TGTGCTCCCAAGGCGCTTCTCTGCGCATGCGCMAAGAGGAGGCTTGGAGCA 260  
 QY 241 GCTGGGGGCTGCGCATCTCTATCTTGGATTAATCCGCTGCGGAGCAGAGCGGA 300  
 DB 261 GCTGGGGGCTGCGCATCTCTATCTTGGATTAATCCGCTGCGGAGCAGAGCGGA 320

QY 301 AGGGGCAAGAGCTTCGTGCGGTGTGGGCCCCCAAGCAGCATCAGAGTGGAGAGTGC 360  
 DB 321 AGGGGCAAGAGCTTCGTGCGGTGTGGGCCCCCAAGCAGCATCAGAGTGGAGAGTGC 380  
 QY 361 AGTCCCGGTGAGTGGCCCTTGGCAGGTCAAGATCACTATGAGAGGCTTCAATGTGTGG 420  
 DB 381 AGTCCCGGTGAGTGGCCCTTGGCAGGTCAAGATCACTATGAGAGGCTTCAATGTGTGG 440  
 QY 421 TGGCTCTCTCGTGTGAGCAGTGGGTGTGTCAGTGTCTCACTGCTTCCCAAGAGCA 480  
 DB 441 TGGCTCTCTCGTGTGAGCAGTGGGTGTGTCAGTGTCTCACTGCTTCCCAAGAGCA 500  
 QY 481 CCACAAGAGAGCTTATGAGGTCAAGCTGGGGCCCAAGCAGTGAATCTCTATCCAGAGA 540  
 DB 501 CCACAAGAGAGCTTATGAGGTCAAGCTGGGGCCCAAGCAGTGAATCTCTATCCAGAGA 560  
 QY 541 CGCCAAAGTCAAGCCTTGAAGAGATATATCCCCCACTACCTTCAGAGAGGCTC 600  
 DB 561 CGCCAAAGTCAAGCCTTGAAGAGATATATCCCCCACTACCTTCAGAGAGGCTC 620  
 QY 601 CCAGGGGCAATTTGCACTCTCCCACTCAGAGAGCCATCACTCTCCCGTACATCCG 660  
 DB 621 CCAGGGGCAATTTGCACTCTCCCACTCAGAGAGCCATCACTCTCCCGTACATCCG 680  
 QY 661 GCCCATCTGCTCCTTGGAGCCCAAGCCTCTTCCCAAGGCTTCACTGCACTGTCAC 720  
 DB 681 GCCCATCTGCTCCTTGGAGCCCAAGCCTCTTCCCAAGGCTTCACTGCACTGTCAC 740  
 QY 721 TGGCTGGGGTCAATGGGCCCCCTCAGTGAAGCTCTTGAAGCCCAAGCCTGACCACT 780  
 DB 741 TGGCTGGGGTCAATGGGCCCCCTCAGTGAAGCTCTTGAAGCCCAAGCCTGACCACT 800  
 QY 781 CGAGGTGCTCGATCAGTGTGAGAGCGTGAATGCTGCTGATCAATCAAGCAGCAGC 840  
 DB 801 CGAGGTGCTCGATCAGTGTGAGAGCGTGAATGCTGCTGATCAATCAAGCAGCAGC 860  
 QY 841 CTGAGAGCCCGCATTTTGTCAAGAGACATGTTGTGTGCTGATGTGAGAGGAGGCA 900  
 DB 860 CTGAGAGCCCGCATTTTGTCAAGAGACATGTTGTGTGCTGATGTGAGAGGAGGCA 919  
 QY 901 AGGAGCTTCCAGAGGTATCTTGGGGGCCCATCTCTCTGCTTGGAGAGGCTCTGTGT 960  
 DB 920 AGGAGCTTCCAGAGGTATCTTGGGGGCCCATCTCTCTGCTTGGAGAGGCTCTGTGT 979  
 QY 961 ACCTGAGGGGATTTGAGCTGGGGAGATGCTGTGGGGGCCCAAGAGGCTGTGTGT 1020  
 DB 980 ACCTGAGGGGATTTGAGCTGGGGAGATGCTGTGGGGGCCCAAGAGGCTGTGTGT 1039  
 QY 1021 ACACTCTGGGCTCCAGTATGCTCTCTGATTCAAAGAGAGGTGACAGAACTCCAGCCTC 1080  
 DB 1040 ACACTCTGGGCTCCAGTATGCTCTCTGATTCAAAGAGAGGTGACAGAACTCCAGCCTC 1099  
 QY 1081 GTGTGTGCTCCCAACCCAGAGAGTCCAGGCCGACAGCAACTCTGTGAGAGCCACTGG 1140  
 DB 1100 GTGTGTGCTCCCAACCCAGAGAGTCCAGGCCGACAGCAACTCTGTGAGAGCCACTGG 1159  
 QY 1141 CCTTCAAGTCTGCGCCAGCCAGGCTTGTGAGAGCCATCTTCTCTGCTCTGGGCC 1200  
 DB 1160 CCTTCAAGTCTGCGCCAGCCAGGCTTGTGAGAGCCATCTTCTCTGCTCTGGGCC 1219  
 QY 1201 TGGCTCTGGGCTCTCTCTCCCATGAGTCTGACGAGACATGAGAGTGGCCCTTCAAGA 1260  
 DB 1220 TGGCTCTGGGCTCTCTCTCCCATGAGTCTGACGAGACATGAGAGTGGCCCTTCAAGA 1279  
 QY 1261 TGGATGATCACTCAAGAGACAGAGAGCTGTGCTCTCTGATGAGCTTTGAGACCAAG 1320  
 DB 1280 TGGATGATCACTCAAGAGACAGAGAGCTGTGCTCTCTGATGAGCTTTGAGACCAAG 1339  
 QY 1321 GCTGACTTGAAGCACTCTCTTTCAGAGACTCTGCGGAGAGGCTGGGGCCCATCTTAT 1380  
 DB 1340 GCTGACTTGAAGCACTCTCTTTCAGAGACTCTGCGGAGAGGCTGGGGCCCATCTTAT 1399  
 QY 1381 CTTTGAAGCCCATCTCTGAGGTGTGCTTTTGGAGCATCACTGAGAGTCAAGATTTA 1440

```

DB 1400 CTTGAGCCCACTTCTTGAGTGTGCTTTTGGAGCAACACAGACAGTCAAGATTTA 1459
QY 1441 CTGCTGTGCAATGAGCCGAGAGCTTGTGCGCCCTCMACACCATGAGACCCCATGGS 1500
DB 1460 CTGCTGTGCAATGAGCCGAGAGCTTGTGCGCCCTCMACACCATGAGACCCCATGGS 1519
QY 1501 CGAGNCTCTGGAGAG--TCTGTGAGCCTTGGATGAAATGAGCCCTGGATTCCACCT 1558
DB 1520 CAGAGCTCTGGAGAGCTCTGAGACCTTGGATGAAATGAGCCCT--GGCTCCACCT 1578
QY 1559 GTTTCCTGGAAGAGCTGCTTCCGCGCCGCTTCCAGACCTGATGAGACATTTTTCG 1618
DB 1579 GTTTCCTGGAAGAGCTG--TCCGCGCCGCTTCCAGACCT--GATGAGACAC--TCTCTGCG 1635
QY 1619 CATTTCCTGCTGTGTTTGGGTGGGCACTTTTGGAGATT 1660
DB 1636 CCTCTCCCTGTGTTCTGGGCTGGGCGCACCTTGTGACAGCTT 1677

RESULT 9
ABL68512
ID ABL68512 standard; DNA; 1834 BP.
XX
AC ABL68512;
XX
DT 15-MAY-2002 (first entry)
XX
DE Kidney cancer related gene sequence SEQ ID NO:6849.
XX
KW Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid;
KM stomach; lung; prostate; pancreas; carcinoma; antitumour; cancerous;
KW cytostatic; gene therapy; antineoplastic; Wilm's tumour; adenocarcinoma;
KW gene; dr.
XX
OS Homo sapiens.
XX
PN MO200194629-A2.
PD
XX 13-DEC-2001.
XX
PF 30-MAY-2001; 2001MO-US10838.
XX
PR 05-JUN-2000; 2000US-209473P.
PR 18-SEP-2000; 2000US-209531P.
PR 18-SEP-2000; 2000US-233135P.
PR 20-SEP-2000; 2000US-234009P.
PR 20-SEP-2000; 2000US-234034P.
PR 20-SEP-2000; 2000US-234052P.
PR 22-SEP-2000; 2000US-234509P.
PR 22-SEP-2000; 2000US-234567P.
PR 25-SEP-2000; 2000US-234924P.
PR 25-SEP-2000; 2000US-234924P.
PR 25-SEP-2000; 2000US-235077P.
PR 25-SEP-2000; 2000US-235082P.
PR 25-SEP-2000; 2000US-235134P.
PR 25-SEP-2000; 2000US-235280P.
PR 26-SEP-2000; 2000US-235637P.
PR 26-SEP-2000; 2000US-235638P.
PR 27-SEP-2000; 2000US-235711P.
PR 27-SEP-2000; 2000US-235720P.
PR 27-SEP-2000; 2000US-235840P.
PR 27-SEP-2000; 2000US-235863P.
PR 28-SEP-2000; 2000US-236028P.
PR 28-SEP-2000; 2000US-236032P.
PR 28-SEP-2000; 2000US-236033P.
PR 28-SEP-2000; 2000US-236034P.
PR 28-SEP-2000; 2000US-236109P.
PR 28-SEP-2000; 2000US-236111P.
PR 29-SEP-2000; 2000US-236842P.
PR 29-SEP-2000; 2000US-236891P.
PR 02-OCT-2000; 2000US-237172P.

```

```

PR 02-OCT-2000; 2000US-237173P.
PR 02-OCT-2000; 2000US-237278P.
PR 02-OCT-2000; 2000US-237294P.
PR 02-OCT-2000; 2000US-237295P.
PR 02-OCT-2000; 2000US-237316P.
PR 03-OCT-2000; 2000US-237425P.
PR 03-OCT-2000; 2000US-237598P.
PR 03-OCT-2000; 2000US-237604P.
PR 03-OCT-2000; 2000US-237606P.
PR 03-OCT-2000; 2000US-237608P.
PR 01-NOV-2000; 2000US-244867P.
PR 01-NOV-2000; 2000US-245084P.
XX
PA (AVAL-) AVALON PHARM.
XX
PI Young PE, Augustus M, Carter KC, Ebner R, Endress G, Horrigan S;
PI Soppet DR, Weaver Z;
XX
DR WPI; 2002-188264/24.
XX
PT Screening for anti-neoplastic agent involves exposing cells to a
PT chemical agent to be tested for anti-neoplastic activity, and
PT determining a change in expression of a gene of a signature gene set -
XX
PS Claim 1; SEQ ID 6849; 44pp; English.
XX
CC The present invention describes a method (M1) for screening for an
CC anti-neoplastic agent. The method involves exposing cells to a chemical
CC agent to be tested for anti-neoplastic activity, determining a change in
CC expression of at least one gene (I) of a signature gene set, where (I)
CC comprises a sequence (S) selected from 8447 sequences (given in ABL6164
CC to ABL70110), or is at least 95% identical to (S), where a change in
CC expression is indicative of anti-neoplastic activity. (I) has cytostatic
CC activity and can be used in gene therapy. M1 can be used for screening
CC an anti-neoplastic agent, and can be used for producing a product which
CC is the data collected with respect to the anti-neoplastic agent as a
CC result of M1, and the data is sufficient to convey the chemical
CC structure and/or properties of the agent. M1 can be used in the
CC treatment of cancer such as colon, breast, stomach, lung, thyroid,
CC oesophageal, ovarian, kidney, prostate or pancreatic cancer,
CC adenocarcinoma, carcinoma, clear cell cancer, infiltrating ductal cancer,
CC infiltrating lobular cancer, squamous cell carcinoma, neuroendocrine
CC carcinoma, papillary carcinoma and Wilm's tumour.
XX
SQ Sequence 1834 BP; 309 A; 621 C; 526 G; 378 T; 0 other;
XX
Query Match 92.2%; Score 1537.4; DB 24; Length 1834;
Best Local Similarity 97.8%; Pred. No. 0;
Matches 1625; Conservative 5; Mismatches 24; Indels 8; Gaps 7;
QY 1 CACACTGCTGCTGCGATATCTCAGAGCTCTCCGTCGAGCGCTCCCTGCTTAGAG 60
DB 22 CACACTGCTGCTGCGATATCTCAGAGCTCTCCGTCGAGCGCTCCCTGCTTAGAG 81
QY 61 GCCAGCTTGAACACTGTGCTGCCCTTTCAGCCGAGATTGAGATCTTCCCTGTAG 120
DB 82 GCCAGCTTGAACACTGTGCTGCCCTTTCAGCCGAGATTGAGATCTTCCCTGTAG 141
QY 121 CCAACATCTGGGTCTGCTGCTTGAACACACCCCAAGCTTCTCACTTCCGTCGAG 180
DB 142 CCAACATCTGGGTCTGCTGCTTGAACACACCCCAAGCTTCTCACTTCCGTCGAG 201
QY 181 TCTGCCCGAGGGGCGCTTGTGCTGGGCGATGAGGAGGGGCTCTGGGGCTGGGA 240
DB 202 TCTGCCCGAGGGGCGCTTGTGCT--GGGCAATGGGCCGAGAGGGGCTCTGGGGCTGGGA 260
QY 241 GCTGGGGGCTGTGGCAATCTGCTTATCTTGGATTACTCCGGTCGAGGACAGAGCGGA 300
DB 261 GCTGGGGGCTGTGGCAATCTGCTTATCTTGGATTACTCCGGTCGAGGACAGAGCGGA 320
QY 301 AGGGGCAAGAGCTCTGCTGGGTGTGCGCCCAAGACGATACAGAGTGGCAGCTGC 360
DB 321 AGGGGCAAGAGCTCTGCTGGGTGTGCGCCCAAGACGATACAGAGTGGCAGCTGC 380

```

QY 361 AGTCGCCGCTCACTGAGCCCTGSCAGGTCAAGCATCACCTATGAAGGCGTCATGTGTGG 420  
 DB 381 AGTCGCCGCTCACTGAGCCCTGSCAGGTCAAGCATCACCTATGAAGGCGTCATGTGTGG 440  
 QY 421 TGGCTCTCTGCTGTCAAGAGAGTGGGTGTGTCAAGCTGTCTCACTGCTTCCCGCAGCA 480  
 DB 441 TGGCTCTCTGCTGTCAAGAGAGTGGGTGTGTCAAGCTGTCTCACTGCTTCCCGCAGCA 500  
 QY 481 CCACAAGAGAGCTTATGAGGTCAAGCTGGGGGGCCCAAGCTAGTACTCTCCACTACCGAGA 540  
 DB 501 CCACAAGAGAGCTTATGAGGTCAAGCTGGGGGGCCCAAGCTAGTACTCTCCACTACCGAGA 560  
 QY 541 CGCAAGGTCAAGACCTTGAAGAGATCATCCCCCAAGCTAGTACTCTCCACTACCGAGA 600  
 DB 561 CGCAAGGTCAAGACCTTGAAGAGATCATCCCCCAAGCTAGTACTCTCCACTACCGAGA 620  
 QY 601 CCAGGGGCGCATTTGACCTCTCTCCACTCAAGAGACCCATCATCTTCTCCGCTACATCCG 660  
 DB 621 CCAGGGGCGCATTTGACCTCTCTCCACTCAAGAGACCCATCATCTTCTCCGCTACATCCG 680  
 QY 661 GCCCATCTCTCTCTCTGAGCCCAAGCGCTCTCTCCCAAGCGGCTCAGCTGAC 720  
 DB 681 GCCCATCTCTCTCTCTGAGCCCAAGCGCTCTCTCCCAAGCGGCTCAGCTGAC 740  
 QY 721 TGGCTGGGGGTCACTGTGGGCCCCCTCACTGAGCCCTCTGAGCCCAAGCCCACTGAGCACT 780  
 DB 741 TGGCTGGGGGTCACTGTGGGCCCCCTCACTGAGCCCTCTGAGCCCAAGCCCACTGAGCACT 800  
 QY 781 CGAGGTGCTCTGATCATGTCTGAGAGCGTGTACTCTCTGTCACATCAGAGCCCAAGC 840  
 DB 801 CGAGGTGCTCTGATCATGTCTGAGAGCGTGTACTCTCTGTCACATCAGAGCCCAAGC 859  
 QY 841 CTGAGAGAGCGGACCTTGTCCAAGAGAGATGTTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 900  
 DB 860 CTGAGAGAGCGGACCTTGTCCAAGAGAGATGTTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 919  
 QY 901 AGGAGCGCTCTGAGGCTGACTCTGTGGGGGCCCACTCTCTGCGCTGTGAGGGGTCTCTGTGT 960  
 DB 920 AGGAGCGCTCTGAGGCTGACTCTGTGGGGGCCCACTCTCTGCGCTGTGAGGGGTCTCTGTGT 979  
 QY 961 ACCTGACGGGCACTTGTGAGCTGGGAGATGCTGTGTGGGGGCCCAAGCGCTGTGTGTGT 1020  
 DB 980 ACCTGACGGGCACTTGTGAGCTGGGAGATGCTGTGTGGGGGCCCAAGCGCTGTGTGTGT 1039  
 QY 1021 AACTCTGCGCTCTGAGCTATGCTCTGTGATCCAAAGAGAGTGAAGAGATCTCCAGCTCTC 1080  
 DB 1040 AACTCTGCGCTCTGAGCTATGCTCTGTGATCCAAAGAGAGTGAAGAGATCTCCAGCTCTC 1099  
 QY 1081 GTGTGGTGGCCCAAAACCCAGAGTCCCAAGCCCAAGCAACTCTGTGAGCAAGCACTGTGT 1140  
 DB 1100 GTGTGGTGGCCCAAAACCCAGAGTCCCAAGCCCAAGCAACTCTGTGAGCAAGCACTGTGT 1159  
 QY 1141 CTTTCAAGCTCTGCCCCAGCCAGGGGCTGTGAGAGCCCATCTTTTCTGCTCTGAGGCC 1200  
 DB 1160 CTTTCAAGCTCTGCCCCAGCCAGGGGCTGTGAGAGCCCATCTTTTCTGCTCTGAGGCC 1219  
 QY 1201 TGGCTCTGGGCTCTCTCTCCATGGCTCAAGAGAGTGAAGTGAAGTGAAGTGAAGTGAAGT 1260  
 DB 1220 TGGCTCTGGGCTCTCTCTCCATGGCTCAAGAGAGTGAAGTGAAGTGAAGTGAAGTGAAGT 1279  
 QY 1261 TGGATGATCACTCACTCAAGAGAGAGCTGTCTCTCTGATGAGCTTGTGAGCCCAAGG 1320  
 DB 1280 TGGATGATCACTCACTCAAGAGAGAGCTGTCTCTCTGATGAGCTTGTGAGCCCAAGG 1339  
 QY 1321 GCGTACCTTGAAGCACTCTTCTCTCAAGAGATCTGTGGGGAGGCTGTGGGGCCCATCTTGTAT 1380  
 DB 1340 GCGTACCTTGAAGCACTCTTCTCTCAAGAGATCTGTGGGGAGGCTGTGGGGCCCATCTTGTAT 1399  
 QY 1381 CTTTGAAGCCATCTTCTTGGGGTGTGCTTTTGGGACATCATCTGAGAGTCAAGAGATTTTAA 1440  
 DB 1400 CTTTGAAGCCATCTTCTTGGGGTGTGCTTTTGGGACATCATCTGAGAGTCAAGAGATTTTAA 1459

QY 1441 CTGCTGTAGCAATGAGCCAGAGCCTCTGAGCCCTCCATMCCACCATGAGCAGCCCATTTGS 1500  
 DB 1460 CTGCTGTAGCAATGAGCCAGAGCCTCTGAGCCCTCCATMCCACCATGAGCAGCCCATTTGS 1519  
 QY 1501 CGAGNTCTTGGGGAG--TCTTGGAGACCTTGTATGAATAATAGCCCTGGGTTCCCACT 1558  
 DB 1520 CGAGNTCTTGGGGAGCCTCTGAGCAGCCCTTGTATGAATAATAGCCCT--GGCTCCCACT 1578  
 QY 1559 GTTTCNAGAGACTGCTTCCGGGCGGCTTCCAGACTNAGATGAGCAATTTTGTTC 1618  
 DB 1579 GTTTCNAGAGACTG--TCCCGGCGGCTTCCAGACT--GATGAGCACA-TCTCTGTGC 1635  
 QY 1619 CATTTCCTGT 1660  
 DB 1636 CTTCTCCCTGT 1677

RESULT 10  
 AAF98698  
 ID AAF98698 standard; DNA; 1796 BP.

AC AAF98698;  
 XX  
 DT 02-JUL-2001 (first entry)  
 XX  
 DE Human ovarian cancer cell expressed sequence 10798.

KW Human; ovarian cancer; identification; detection; characterisation;  
 XX tumour; kinase; marker; cytosolic; antisense gene therapy; ds.  
 OS Homo sapiens.

PN WO200118542-A2.

PD 15-MAR-2001.

PF 01-SEP-2000; 2000MO-US24199.

PR 03-SEP-1999; 99US-0152547.

PR 16-MAR-2000; 2000US-0190347.

PR 21-MAR-2000; 2000US-0191321.

PR 31-MAY-2000; 2000US-0208382.

PR 20-JUL-2000; 2000US-0220467.

PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.

PI Lee J, Thompson P, Lillie J;

DR WPI; 2001-211428/21.

PT Detection, assessment, prevention and therapy of ovarian cancer.

PT comprises detecting changes in the expression of a variety of markers -

PS Claim 1; Page 1001-1002; 1198pp; English.

CC The present invention describes a method for assessing whether a patient

CC is afflicted with ovarian cancer by comparing: (1) the expression of a

CC marker (1) (see AAF98698 to AAF98730), in a patient sample; and (2) the

CC normal level of expression of (1) in a control non-ovarian cancer

CC sample, where a significant difference between the level of expression

CC in (a) and (b) is an indication that the patient is afflicted with

CC ovarian cancer. (1) have cytosolic activities and can be used in

CC antisense gene therapy. The method, compositions and kits from the

CC present invention can be used for: (1) assessing and treating ovarian

CC cancer; (2) making isolated hybridoma, which produces an antibody useful

CC for ovarian cancer assessment; and (3) inhibiting ovarian cancer in a

CC patient. AAF98573 to AAF98593 represent human kinase marker primers and

CC probes which are used in the exemplification of the present invention.

SQ Sequence 1796 BP; 327 A; 599 C; 511 G; 359 T; 0 other;

Query Match 88.5%; Score 1475.8; DB 22; Length 1796;  
 Best Local Similarity 97.7%; Pred. No. 1.1e-307;

Matches	1553;	Conservative	5;	Mismatches	25;	Indels	7;	Gaps	6;
QY	73	CACTTGTGCCCCCTTTTCCAGCCGGAATTCGTGATCTTCTCTGTGAGCAACATCTGGG							132
Db	16	CGCTTGCCGCCCCCTTTTCCAGCCGGAATTCGTGATCTTCTCTGTGAGCAACATCTGGG							75
QY	133	TCCTGACCTTTCGACACACCCCAAGGCTTCTACTTGTGCGCTGGAGCTGCCCCAGGG							192
Db	76	TCCTGACCTTTCGACACACCCCAAGGCTTCTACTTGTGCGCTGGAGCTGCCCCAGGG							135
QY	193	GCCCTTGTCTGGGCGCATGGCCMAGAGGGGCTCTGGGGCCTGGGCACTGGGGCTGT							252
Db	136	GCCCTTGTCTGGGCGCATGGCCMAGAGGGGCTCTGGGGCCTGGGCACTGGGGCTGT							195
QY	253	GGCCATTCTGCTCTATCTTGTGATTAATCCGGTGGGGGACAGAGAGCCGAGGGGAGAAC							312
Db	196	GGCCATTCTGCTCTATCTTGTGATTAATCCGGTGGGGGACAGAGAGCCGAGGGGAGAAC							255
QY	313	TYCCCTGGGCTGGGGCCCCCAAGACAGCATCAAGGTGGACAGAGTGCAGCCGGTCA							372
Db	256	TYCCCTGGGCTGGGGCCCCCAAGACAGCATCAAGGTGGACAGAGTGCAGCCGGTCA							315
QY	373	GTGGCCCTGGGACAGTCAGATCACTATGAGGCGTCAATGTGTGTGTGTGTGTCTCTGT							432
Db	316	GTGGCCCTGGGACAGTCAGATCACTATGAGGCGTCAATGTGTGTGTGTGTGTCTCTGT							375
QY	433	GTCTGAGCAGTGGGTCTGTCACTGTCTCACTGTCTTCCCGAGGACACACAGGAAGC							492
Db	376	GTCTGAGCAGTGGGTCTGTCACTGTCTCACTGTCTTCCCGAGGACACACAGGAAGC							435
QY	493	CTATGAGTCAGAGCTGGGGCCCCACAGGCTAGACTCTACTCCGAGAGACGCCAAGGTCA							552
Db	436	CTATGAGTCAGAGCTGGGGCCCCACAGGCTAGACTCTACTCCGAGAGACGCCAAGGTCA							495
QY	553	CACCCGAAAGACATATCCCCCAACCCCAAGCTACTCCAGAGAGGCTCCAGGGGCAAT							612
Db	496	CACCCGAAAGACATATCCCCCAACCCCAAGCTACTCCAGAGAGGCTCCAGGGGCAAT							555
QY	613	TGCACTCTTCCAACTGACAGACCCCATACCTTCTCCGCTACATCCGGCCCATCTGGCT							672
Db	556	TGCACTCTTCCAACTGACAGACCCCATACCTTCTCCGCTACATCCGGCCCATCTGGCT							615
QY	673	CCCTGAGCCAAAGGCTCTTCCCAAGGCTCCACTGACCTGTCACTGGCTGGGGTCA							732
Db	616	CCCTGAGCCAAAGGCTCTTCCCAAGGCTCCACTGACCTGTCACTGGCTGGGGTCA							675
QY	733	TGTGGCCCCCTCACTGAGCTCTCTGACGCCCAAGCACTGACAGAACTCGAGGTCTCT							792
Db	676	TGTGGCCCCCTCACTGAGCTCTCTGACGCCCAAGCACTGACAGAACTCGAGGTCTCT							735
QY	793	GATCAGTCGTGAGACGTGTAACTGCTGTAAACAATCGACGCCAGCTGAGAGCCGC							852
Db	736	GATCAGTCGTGAGACGTGTAACTGCTGTAAACAATCGACGCCAGCTGAGAGCCGC							794
QY	853	ACTTTTGCAGAGAGACATGATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT							912
Db	795	ACTTTTGCAGAGAGACATGATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT							854
QY	913	AGGGTGACTCTGGGGGCCCACTCTCTGCTGCTGTGAGAGGCTCTGTGTAACCTGACGGA							972
Db	855	AGGGTGACTCTGGGGGCCCACTCTCTGCTGCTGTGAGAGGCTCTGTGTAACCTGACGGA							914
QY	973	TGTGAGCTGGGAGATGCTGTGTGGGCCCGCAACAGGCTGTGTGTAACTCTGGCCT							1032
Db	915	TGTGAGCTGGGAGATGCTGTGTGGGCCCGCAACAGGCTGTGTGTAACTCTGGCCT							974
QY	1033	CCAGCTATGCTCTCTGATCCAAAGAGAGAGAACTCCAGGCTCGTGTGTGTGTGTGT							1092
Db	975	CCAGCTATGCTCTCTGATCCAAAGAGAGAGAACTCCAGGCTCGTGTGTGTGTGTGT							1034
QY	1093	AAACCCAGAGTCCACGCCGACAGCAACCTGTGTGACGACCACTTGACCTCTG							1152
Db	1035	AAACCCAGAGTCCACGCCGACAGCAACCTGTGTGACGACCACTTGACCTCTG							1094

QY	1153	CCCCAGCCAGGGCTTGTGAGGCCCATCTTTTCTGCTCTGAGGCTGGGCTGTGGCC	1212
Db	1095	CCCCAGCCAGGGCTTGTGAGGCCCATCTTTTCTGCTCTGAGGCTGGGCTGTGGCC	1154
QY	1213	TCCTTCCCATAGGCTGACGAGACATGAGCTGGGCTTACTTCCAGATGATGATCAC	1272
Db	1155	TCCTTCCCATAGGCTGACGAGACATGAGCTGGGCTTACTTCCAGATGATGATCAC	1214
QY	1273	ACTCAAGACAGAGACCTGTGCTTCCCTGATAGGCTTTGGACCCAGGGGCTGACTTGA	1332
Db	1215	ACTCAAGACAGAGACCTGTGCTTCCCTGATAGGCTTTGGACCCAGGGGCTGACTTGA	1274
QY	1333	CCACTCTTCTCTTCAAGACTGTGCGGAGGCTGGGGCCCATCTTGTATCTTGAAGCCAT	1392
Db	1275	CCACTCTTCTCTTCAAGACTGTGCGGAGGCTGGGGCCCATCTTGTATCTTGAAGCCAT	1334
QY	1393	TCCTTGGGTGTGCTTTTGGACCATGACATGAGTCAAGAGTCTTACTGCTGTAGCA	1452
Db	1335	TCCTTGGGTGTGCTTTTGGACCATGACATGAGTCAAGAGTCTTACTGCTGTAGCA	1394
QY	1453	ATGGCCAGAGCTCTGGGCCCTCAMCCACCATGAGCCAGCCATTGGGCGAGNTCCTGGG	1512
Db	1395	ATGGCCAGAGCTCTGGGCCCTCAMCCACCATGAGCCAGCCATTGGGCGAGNTCCTGGG	1454
QY	1513	GAG--TCCGGGACCTTGGGTATGAAATGAGCCCTGGGTTCCACCTGTTCTGAGAGA	1570
Db	1455	GAGCTTCTGGGACCTTGGGCTATGAAATGAGCCCT--GGCTCCACCTGTTCTGAGAGA	1513
QY	1571	CTGCTTCCCGGCCCGCTTCCCAAGCTGATGATGAGCAATTTTTCCTGCTGCTGG	1630
Db	1514	CTGC--TCCGGGCCCGGCTGCCAGACT--GATAGAGACA--TCCTCTGCTGCTGCTGG	1570
QY	1631	TTTTTGGGTGGGCACTTTTGGAACTTTT	1660
Db	1571	TTCTGGGCTGGGCACTTTGTGAGCTT	1600

## RESULT 11

ABT31936

ID ABT31936 standard; DNA; 1733 BP.

XX ABT31936;

XX 01-MAY-2003 (first entry)

XX Human breast cancer / ovarian cancer related coding sequence #43.

XX Human; gene; de; cytosolic; breast cancer; ovarian cancer.

XX Homo sapiens.

XX WO2003000012-A2.

XX 03-JAN-2003.

XX 21-JUN-2002; 2002MO-US19773.

XX 21-JUN-2001; 2001US-300159P.

XX 27-JUN-2001; 2001US-301351P.

XX (MILL-) MILLENNIUM PHARM INC.

XX Veibby OP;

XX WPI; 2003-267848/26.

XX P-PSDB; ABJ37067.

XX Determining the presence of breast cancer in an individual, involves

XX using specific polynucleotide markers -

XX Disclosure; Page 192-193; 233pp; English.

XX PS

CC The invention comprises a method for assessing whether a patient is  
 CC afflicted with breast cancer or ovarian cancer. The method involves the  
 CC use of specific DNA markers. The method of the invention is useful in the  
 CC detection and treatment of ovarian and breast cancer. DNA sequences  
 CC ABT31894 - ABT31949 encode human breast/ovarian cancer-related proteins.

XX  
 SQ Sequence 1733 BP, 311 A, 578 C, 500 G, 344 T, 0 other;

Query Match 85.2%; Score 1420.8; DB 25; Length 1733;  
 Best Local Similarity 97.6%; Pred. No. 7,1e-296;  
 Matches 1498; Conservative 5; Mismatches 25; Indels 7; Gaps 6;

QY 128 CTGGGCTCTGCTTCCGACACCAACCCCAAGCTTCTTAACCTTGGTGCCTGAGCTTCC 187  
 DB 18 CTGGGCTCTGCTTCCGACACCAACCCCAAGCTTCTTAACCTTGGTGCCTGAGCTTCC 77  
 QY 188 CAGGGGCTCTGCTTCCGACACCAACCCCAAGCTTCTTAACCTTGGTGCCTGAGCTTCC 247  
 DB 78 CAGGGGCTCTGCTTCCGACACCAACCCCAAGCTTCTTAACCTTGGTGCCTGAGCTTCC 137  
 QY 248 GCTGTGGCCATTTCTGCTTATCTTGGATTACTCCGGTCGGGAGCAGAGCGGAAAGGAGCA 307  
 DB 138 GCTGTGGCCATTTCTGCTTATCTTGGATTACTCCGGTCGGGAGCAGAGCGGAAAGGAGCA 197  
 QY 308 GAAAGCTTCTGCGGTGGCCGCCCAAGCAGCATCAAGTGGCAGAGTGCAGTCCGCC 367  
 DB 198 GAAAGCTTCTGCGGTGGCCGCCCAAGCAGCATCAAGTGGCAGAGTGCAGTCCGCC 257  
 QY 368 GGTCAATGGCCCTTGGGAGGTCAAGCATCACTTAAGAGGGGTCCATGTGTGTGGTCTCT 427  
 DB 258 GGTCAATGGCCCTTGGGAGGTCAAGCATCACTTAAGAGGGGTCCATGTGTGTGGTCTCT 317  
 QY 428 CTGTGTCTGAGCAGTGGGT 487  
 DB 318 CTGTGTCTGAGCAGTGGGT 377  
 QY 488 GAAAGCTTATAGGTCAAGCTGGGGGGCCCAAGCAGTGAATCTCTTCCAGAGCGCCAG 547  
 DB 378 GAAAGCTTATAGGTCAAGCTGGGGGGCCCAAGCAGTGAATCTCTTCCAGAGCGCCAG 437  
 QY 548 GTTCAGACCTTGAAGGACATATCCCCCAAGCTCACTCTTCCAGAGCGGCTTCCAGGCG 607  
 DB 438 GTTCAGACCTTGAAGGACATATCCCCCAAGCTCACTCTTCCAGAGCGGCTTCCAGGCG 497  
 QY 608 GACATTGCACTCTCTCAACTCAGCAGACCATCACTCTCTCCGCTACATCCGGGCCCATC 667  
 DB 498 GACATTGCACTCTCTCAACTCAGCAGACCATCACTCTCTCCGCTACATCCGGGCCCATC 557  
 QY 668 TGCCTTCCCTGAGCCCAAGCGCTTCTTCCCAAGCGGCTTCCACTGCACTGTGCTG 727  
 DB 558 TGCCTTCCCTGAGCCCAAGCGCTTCTTCCCAAGCGGCTTCCACTGCACTGTGCTG 617  
 QY 728 GGTCAATGGCCCTTCCGACACCAACCCCAAGCTTCTTAACCTTGGTGCCTGAGCTTCC 787  
 DB 618 GGTCAATGGCCCTTCCGACACCAACCCCAAGCTTCTTAACCTTGGTGCCTGAGCTTCC 677  
 QY 788 CCTGTATGATGCTGAGAGCGTGTAACTGCTTCAACATGAGAGCCCAAGCTTGA3A 847  
 DB 678 CCTGTATGATGCTGAGAGCGTGTAACTGCTTCAACATGAGAGCCCAAGCTTGA3A 736  
 QY 848 GCGGCACTTTGTTCAGAGAGACATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 907  
 DB 737 GCGGCACTTTGTTCAGAGAGACATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 796  
 QY 908 CTGCGAGGCTGACTCTGGGGGCCCACTCTCTGCTGTGAGAGGGTCTTGGTAACTTAAC 967  
 DB 797 CTGCGAGGCTGACTCTGGGGGCCCACTCTCTGCTGTGAGAGGGTCTTGGTAACTTAAC 856  
 QY 968 GGGCATTTGTGAGCTGGGAGATGCTGTGGGGGCCCAAGCAGGCTGTGTGTGTGTGTGTGT 1027  
 DB 857 GGGCATTTGTGAGCTGGGAGATGCTGTGGGGGCCCAAGCAGGCTGTGTGTGTGTGTGTGT 916  
 QY 1028 GGCCTTCAGCTATGCTCTGTGATCCAAAGGATGACAGAACTTCAGCCTTGTGTGTGT 1087

DB 917 GGCCTTCAGCTATGCTCTCTGTGATCCAAAGGATGACAGAACTTCAGCCTTGTGTGTGT 976  
 QY 1088 GGGCCAAACCCAGAGGTCCAGCCCGAGCAGCAACCTCTGTGACAGCACTTGGCCCTTCAG 1147  
 DB 977 GGGCCAAACCCAGAGGTCCAGCCCGAGCAGCAACCTCTGTGACAGCACTTGGCCCTTCAG 1036  
 QY 1148 CTCTGCCCCAGGCCCAAGGCTTGTGAGGCCCATCTTCTTCTGCTTGGGCTTGGCTCT 1207  
 DB 1037 CTCTGCCCCAGGCCCAAGGCTTGTGAGGCCCATCTTCTTCTGCTTGGGCTTGGCTCT 1096  
 QY 1208 GGGCCTCTCTTCCCAATGAGCTCAGCAGAGCACTGAGCTGAGCCCTTACTTCCAGATGATGC 1267  
 DB 1097 GGGCCTCTCTTCCCAATGAGCTCAGCAGAGCACTGAGCTGAGCCCTTACTTCCAGATGATGC 1156  
 QY 1268 ATCAGACTCAAGAGAGAGAGGCTGTGCTTCCCTGATAGGCCCTTTGAGACCCAGAGGCTGAC 1327  
 DB 1157 ATCAGACTCAAGAGAGAGGCTGTGCTTCCCTGATAGGCCCTTTGAGACCCAGAGGCTGAC 1216  
 QY 1328 TTGAGCACTCTTCTTCTTCTGAGACTCTGCGGAGGCTGGGGCCCACTTGTGATCTTTAG 1387  
 DB 1217 TTGAGCACTCTTCTTCTTCTGAGACTCTGCGGAGGCTGGGGCCCACTTGTGATCTTTAG 1276  
 QY 1388 CCCATTCTTCTGAGGTGTGCTTTTGGACCATCACTGAGAGTCAAGAGTTTAACTGCTG 1447  
 DB 1277 CCCATTCTTCTGAGGTGTGCTTTTGGACCATCACTGAGAGTCAAGAGTTTAACTGCTG 1336  
 QY 1448 TAGCAATGGCCAGAGCTTGTGCCCCCTCMCCACCATAGACCAAGCCCATTTGGSCAGAGTTC 1507  
 DB 1337 TAGCAATGGCCAGAGCTTGTGCCCCCTCMCCACCATAGACCAAGCCCATTTGGSCAGAGTTC 1396  
 QY 1508 CTGGGGAG--TCTTGGGACCTTGTGTATGAATAAGAGCCTGGGTTCACCTGTTTCTN 1565  
 DB 1397 CTGGGGAGCTCTTGGGAGCCTTGTGCCCCCTCMCCACCATAGACCAAGCCCATTTGGSCAGAGTTC 1455  
 QY 1566 GAAAGCTGCTTCCGCGCCGCTTCCAGACTGATGAGACATTTTTCCTTTCC 1625  
 DB 1456 GAAAGCTGCTTCCGCGCCGCTTCCAGACTGATGAGACATTTTTCCTTTCC 1512  
 QY 1626 CTGTGTTTTGGGTTGGGCACTTTTGTGAACTTT 1660  
 DB 1513 CTGTGTTTTGGGTTGGGCACTTTTGTGCACTTT 1547

RESULT 12  
 ABN85392  
 ID ABN85392 standard; DNA, 1726 BP.  
 XX  
 AC ABN85392;  
 XX  
 DT 21-OCT-2002 (first entry)  
 XX  
 DE Human NOV14a, prostaasin-like protein, coding sequence.  
 XX  
 XX Human; NOV14a; cytostratic; Cardiant; Antiinflammatory; Immunosuppressive;  
 KW Antiallergic; Haemostatic; Anti-HIV; Antidiabetic; Anorectic;  
 KW Antibacterial; Nephroretropic; Hepatotropic; Neuroprotective; Noctropic;  
 KW Gene Therapy; NOV; cancer; heart disease; Inflamation;  
 KW autoimmune disorder; allergy; blood disorder; AIDS; diabetes;  
 KW obesity; asthma; Iga nephropathy; cirrhosis; arthritis;  
 KW Alzheimer's disease; infection; stroke; muscular dystrophy; epilepsy;  
 KW wasting disorder; prostaasin-like protein; chromosome 16p11.2;  
 KW gene; de.  
 XX  
 OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 229..1152  
 FT /tag=a  
 FT /product="NOV14a"  
 XX  
 PN WO200255704-A2.

XX 18-JUL-2002.  
 PD  
 XX  
 PF 09-JAN-2002; 2002MO-US00554.  
 XX  
 XX 09-JAN-2001; 2001US-260417P.  
 PR 10-JAN-2001; 2001US-260831P.  
 PR 28-FEB-2001; 2001US-272338P.  
 PR 09-MAR-2001; 2001US-274876P.  
 PR 18-APR-2001; 2001US-284704P.  
 XX  
 PA (CURA-) CURAGEN CORP.  
 XX  
 PI Padigaru M, Li L, Zerhusen BD, Casman SJ, Shenoy S, Spytek KA;  
 PI Zhong M, Gangoli EA, Burgess CE, Patturajan M, Vernet CM;  
 PI Taylor S, Tcheney VT, Miller CE, Guo X, Boldog FL, Grose WM;  
 PI Alsobrook JF, Gerlach V, Edinger S, Rothenberg ME, Ellerman K;  
 PI MacDougall J, Malyankar U, Millet I, Peyman J, Smithson G;  
 PI Gunther E, Stone DJ;  
 XX  
 DR WPI; 2002-590674/63.  
 P-PSDB; ABB98415.  
 XX  
 XX NOVA polypeptides and encoding polynucleotides, useful for preventing  
 PT or treating NOVA-associated disorders e.g. cancer, inflammation, or  
 PT Alzheimer's disease, and in chromosome mapping, tissue typing or  
 PT pharmacogenomics -  
 XX  
 PS Claim 9; Page 97-98; 358pp; English.  
 XX  
 XX The present sequence is a coding sequence for a NOV protein. The  
 CC NOV proteins and coding sequences are useful for treating or preventing  
 CC NOV-associated disorders or in the manufacture of a medicament for  
 CC treating the disorders, such as cancer, heart disease, inflammation,  
 CC autoimmune disorders, allergies, blood disorders, AIDS, diabetes,  
 CC obesity, asthma, IGA nephropathy, cirrhosis, arthritis, Alzheimer's  
 CC disease, infections (e.g. bacterial, viral, parasitic), stroke, muscular  
 CC dystrophy, epilepsy, and other wasting disorders associated with chronic  
 CC diseases. NOV14a is a prostatic-like protein, and the NOV14a coding  
 CC sequence localises to chromosome 16p11.2.  
 XX  
 XX Sequence 1726 BP; 284 A; 581 C; 499 G; 362 T; 0 other;  
 SQ  
 Query Match 78.2%; Score 1305; DB 24; Length 1726;  
 Best Local Similarity 91.0%; Pred. No. 5.5e-271;  
 Matches 1513; Conservative 5; Mismatches 28; Indels 116; Gaps 8;

DB 381 AGTGGCCGGTCAATGAGCCCTGGACAGTCAATCACTAATGAAGCGCTCATGTGTGG 440  
 QY 421 TGGCTCTCTGATGCTAGACAGTGGGTGCTGACGTGCTCACTGTTCCCGACGAGA 480  
 DB 441 TGGCTCTCTGATGCTAGACAGTGGGTGCTGACGTGCTCACTGTTCCCGACGAGA 500  
 QY 481 CCACAGGAAGCCTATAGAGTCAAGCTGGGGGCCACAGCTAGACTCTACTCCGAGGA 540  
 DB 501 CCACA----- 505  
 QY 541 CGCAAGTCAAGCCTTAAGACATCATCCCAAGCCAGCTACTCCAGAGAGGCTC 600  
 DB 506 -----AGGCTC 512  
 QY 601 CCAGGCGCATTTGACACTCTCCCACTAGACAGACCATCACTTCCCGCTACATCCG 660  
 DB 513 CCAGGCGCATTTGACACTCTCCCACTAGACAGACCATCACTTCCCGCTACATCCG 572  
 QY 661 GCCCATCTGCTCTCCCTGACGCAAGCGCTCTTCCCAAGCGCTCCACTGCACTGTAC 720  
 DB 573 GCCCATCTGCTCTCCCTGACGCAAGCGCTCTTCCCAAGCGCTCCACTGCACTGTAC 632  
 QY 721 TGGCTGGGTCAATGTGGCCCTCCCTCACTAGACCTCTCCCAAGCCCACTGAGAACT 780  
 DB 633 TGGCTGGGTCAATGTGGCCCTCCCTCACTAGACCTCTCCCAAGCCCACTGAGAACT 692  
 QY 781 CGAGTGGCTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 840  
 DB 693 CGAGTGGCTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 751  
 QY 841 CTGAGAGCCGCACTTTGTCAGAGACATGATGATGATGATGATGATGATGATGATG 900  
 DB 752 CTGAGAGCCGCACTTTGTCAGAGACATGATGATGATGATGATGATGATGATGATG 811  
 QY 901 AGAGCCCTGCAAGGTGATCTGAGGAGCCCACTCTCCCTGAGAGGCTCTGTGT 960  
 DB 812 AGAGCCCTGCAAGGTGATCTGAGGAGCCCACTCTCTCTGAGAGGCTCTGTGT 871  
 QY 961 ACCTGAGGAGCATTTGAGAGTGGAGAGATGCTGTGGGAGCCGCAAGAGGCTGTGT 1020  
 DB 872 ACCTGAGGAGCATTTGAGAGTGGAGAGATGCTGTGGGAGCCGCAAGAGGCTGTGT 931  
 QY 1021 ACACTGTGCTCTGACATATGCTCTGTGATCCAAAGAGTACAGAACTCCAGCTTC 1080  
 DB 932 ACACTGTGCTCTGACATATGCTCTGTGATCCAAAGAGTACAGAACTCCAGCTTC 991  
 QY 1081 GTGTGTGCCCCCAACCCAGAGTCCAGCCGCAAGCAACTCTGTGAGAGCACTGTG 1140  
 DB 992 GTGTGTGCCCCCAACCCAGAGTCCAGCCGCAAGCAACTCTGTGAGAGCACTGTG 1051  
 QY 1141 CCTTGAAGCTTGGCCGAGCCAGAGGCTGTGAGGAGCCCAATCTTCTGCTGTGGCC 1200  
 DB 1052 CCTTGAAGCTTGGCCGAGCCAGAGGCTGTGAGGAGCCCAATCTTCTGCTGTGGCC 1111  
 QY 1201 TGGCTGTGGGCTCTCTCTCCCAAGTCTGAGAGCACTGAGCTGAGCTTCACTTCA 1260  
 DB 1112 TGGCTGTGGGCTCTCTCTCCCAAGTCTGAGAGCACTGAGCTGAGCTTCACTTCA 1171  
 QY 1261 TGGATGATCACTCAAGAGCAGAGGCTGTGCTTCTCTGATGAGCTTGTGAGCCAG 1320  
 DB 1172 TGGATGATCACTCAAGAGCAGAGGCTGTGCTTCTCTGATGAGCTTGTGAGCCAG 1231  
 QY 1321 GCTGATCTGAGCACTCTCTTCAAGACTGTGGAGAGGCTGTGGAGCCCACTTGTAT 1380  
 DB 1232 GCTGATCTGAGCACTCTCTTCAAGACTGTGGAGAGGCTGTGGAGCCCACTTGTAT 1291  
 QY 1381 CTTTGAAGCCATCTTCTGTGGTGTCTTTTGGAGCAATCATGAGAGTCAAGAGTTTA 1440  
 DB 1292 CTTTGAAGCCATCTTCTGTGGTGTCTTTTGGAGCAATCATGAGAGTCAAGAGTTTA 1351  
 QY 1441 CTGCTGTAGCAATGAGCCAGAGCTTGTGGCCCTTCAACCACTGAGACAGCCATTGG 1500  
 DB 1352 CTGCTGTAGCAATGAGCCAGAGCTTGTGGCCCTTCAACCACTGAGACAGCCATTGG 1411



QY 1501 CGAGTCTCTGGGAG--TCCTGGACCTTGGYATGAAAAGACCCCTGGGTTCCCACT 1558  
 DB 1412 CGAGCTCTGGGAGGAGCTCTGGAGACCTTGCTATGAAATGAGCCCT-GGCTCCCACT 1470  
 QY 1559 GTTTCNMGAAAGACTGCTTCCCGCCGCTTCCCACTGATGAGACACTTTTTCG 1618  
 DB 1471 GTTTCGGAAGACTGC-TCCCGGCCCGCCCTGCCAGACT-GATGAGACCA-TCTCTCTGC 1527  
 QY 1619 CATTTCCTGTGTGTTTGGGTGGGCACTTTTGAAGTTT 1660  
 DB 1528 CTTCTCCTGTGTCTTCTGGGCTGGGCGCACCTTTGTGCACTT 1569  
 RESULT 13  
 ABZ80059  
 ID ABZ80059 standard; DNA; 2131 BP.  
 XX ABZ80059;  
 AC  
 XX 21-MAY-2003 (first entry)  
 XX  
 XX  
 DE Human AGE receptor binding related nucleotide sequence SEQ ID NO:61.  
 XX  
 XX Human; signal transfer; advanced glycation end product receptor;  
 KW AGE receptor; RAGE; antidiabetic; neurotrophic; neuroprotective;  
 KW cytostatic; antiinflammatory; gene therapy; diabetes; cancer;  
 KW diabetic complication; Alzheimer's disease; amyloidosis;  
 KW periodontal disease; gene; ds.  
 XX  
 XX Homo sapiens.  
 OS  
 XX WO2003008446-A1.  
 PN  
 XX 30-JAN-2003.  
 PD  
 XX 18-JUL-2002; 2002WO-IP07344.  
 PF  
 XX 19-JUL-2001; 2001JP-0219122.  
 PR  
 XX (MITS-) MITSUBISHI PHARMA CORP.  
 PA  
 XX Tsuchida J;  
 PI  
 XX WPI; 2003-239310/23.  
 DR  
 XX Polypeptide binding to the cytoplasmic domain of advanced glycation  
 PT end product receptor for treatment of cancer and diabetes  
 XX  
 XX Claim 6; Page 77-78; 104pp; Japanese.  
 PS  
 XX  
 XX  
 CC The present invention describes polypeptides (I) which bind to the  
 CC cytoplasmic domain of the advanced glycation end-product (AGE) receptor  
 CC (also known as RAGE), either directly or indirectly, and inhibit the  
 CC signal transduction from the binding of RAGE with its ligand to the  
 CC activation of nuclear factor kappa beta (NFkB). Also described:  
 CC (1) polynucleotides encoding (I); (2) expression vectors containing  
 CC polynucleotides encoding (I); (3) microorganisms and cells transformed  
 CC by these vectors; (4) preparation of (I) by culture of the transformants;  
 CC (5) gene therapy drug compositions containing vectors for expression of  
 CC (I) in animals; (6) screening compounds inhibiting or promoting the  
 CC binding of (I) to the cytoplasmic domain of the RAGE, or modifying the  
 CC expression of (I); (7) compounds identified by this screening method;  
 CC (8) drug compositions containing the compounds identified by the  
 CC screening method; (9) antibodies recognizing all or part of (I);  
 CC (10) polynucleotides at least 20 bases in length, hybridizing to all or  
 CC part of the polynucleotides encoding (I); and (11) probes containing  
 CC these polynucleotides, and diagnostic reagents containing them. (I) have  
 CC antidiabetic, neurotrophic, neuroprotective, cytostatic and antiinflammatory  
 CC activities, and can be used for the regulation of AGE signal transduction  
 CC into the cell via RAGE and for gene therapy. AGE are sugar reaction  
 CC products with amine groups which accumulate in association with diabetes,  
 CC cancer and other diseases. (I) can be used for the treatment and

CC prevention of diabetes, diabetic complications, Alzheimer's disease,  
 CC cancer, amyloidosis and periodontal disease, by conventional or gene  
 CC therapy. ABZ80031 to ABZ80071 and ABP96322 to ABP96373 represent  
 CC sequences given in the exemplification of the present invention.

XX Sequence 2131 BP; 406 A; 645 C; 642 G; 438 T; 0 other;

Query Match 52.7%; Score 879.2; DB 25; Length 2131;  
 Best Local Similarity 84.2%; Pred. No. 1.8e-179;  
 Matches 1158; Conservative 3; Mismatches 29; Indels 185; Gaps 8;

QY 472 CAGCGACACCAAGAAAGCTATGAGTCAAGCTGGGGGCCACAGCTAGACTCTTA 531  
 DB 618 CAGCGACACCAAGAAAGCTATGAGTCAAGCTGGGGGCCACAGCTAGACTCTTA 677  
 QY 532 CTCGAGGAGCGGCAAGGTGACACCTGAAAGACATCATCCCAACCCAGTACTTCA 591  
 DB 678 CTCGAGGAGCGGCAAGGTGACACCTGAAAGACATCATCCCAACCCAGTACTTCA 737  
 QY 592 GGAGGGCTCCAGGGCGGACATGACCTCTCCAGCTCAGCAGACCATCACCCTCTCCG 651  
 DB 738 GGAGGGCTCCAGGGCGGACATGACCTCTCCAGCTCAGCAGACCATCACCCTCTCCG 797  
 QY 652 CTACATCCGGGCCATCTGCTCTCTGACAGCCAAAGCTTCTCCCAAGGCTCCACTG 711  
 DB 798 CTACATCCGGGCCATCTGCTCTCTGACAGCCAAAGCTTCTCCCAAGGCTCCACTG 857  
 QY 712 CACTGTCACTGGCTGGGGTCAATGTGGCCCTT----- 743  
 DB 858 CACTGTCACTGGCTGGGGTCAATGTGGCCCTTCAAGTGAAGTGGAGCTGGTGCCTAGA 917  
 QY 744 -----CAG 746  
 DB 918 GGTGTGAGGGGCGACCTGACTTGGGGAAGGGCCAGGGTAAGCTCTTTTACCCCAAG 977  
 QY 747 TGAGCTCTCTAGCGCCCAAGCCACTGACAGCACTGAGGTGCTTGTATCACTGTGAGA 806  
 DB 978 TGAGCTCTCTAGCGCCCAAGCCACTGACAGCACTGAGGTGCTTGTATCACTGTGAGA 1037  
 QY 807 CGTGTAACTGCTTGTATCAACAATGACAGCCCAAGCTGAGAGAGCCGACTTGTCCAAG 866  
 DB 1038 CGT-GTAACTGCTTGTATCAACAATGACAGCCCAAGCTGAGAGAGCCGACTTGTCCAAG 1096  
 QY 867 GACATGATGTGTGCTGAGGCTATGTGAGGGGGGCAAGAGCGCTGC----- 911  
 DB 1097 GACATGATGTGTGCTGAGGCTATGTGAGGGGGGCAAGAGCGCTGCACAGTAAGCACAGGC 1156  
 QY 912 ----- 911  
 DB 1157 CCGGGGGGAGATGACCAAGTGCAACTTGGAAAGAGGGCTGGCCGGTCTGATGGCTGC 1216  
 QY 912 -----CAGGTGACTTGGGGGCGCCACTCTCTGCTGCTGTGAGGGTC 954  
 DB 1217 TGTGTGCTTCTCTCTTCAAGGTGACTTGGGGGCGCCACTCTCTGCTGCTGTGAGGGTC 1276  
 QY 955 TGTGTGACTTCAAGGGCAATGTGAGCTGGGGGAGATGCTGTGGGGCCCGCAACAGGCTG 1014  
 DB 1277 TGTGTGACTTCAAGGGCAATGTGAGCTGGGGGAGATGCTGTGGGGCCCGCAACAGGCTG 1336  
 QY 1015 GTGTGACACTTGTGCTTCAGACTATGCTCTGTGATCAAAAGAGGTGACAGAACTCC 1074  
 DB 1337 GTGTGACACTTGTGCTTCAGACTATGCTCTGTGATCAAAAGAGGTGACAGAACTCC 1396  
 QY 1075 AGCTCTGTGTGTGCTCCCAAAACCCAGAGTCCAGCCCGACAGCAACTCTGTGAGCC 1134  
 DB 1397 AGCTCTGTGTGTGCTCCCAAAACCCAGAGTCCAGCCCGACAGCAACTCTGTGAGCC 1456  
 QY 1135 AACTGGCTTGAAGCTCTGCCAGCGGCTTGTGAGGCGCACTTTTCTGCGCTC 1194  
 DB 1457 AACTGGCTTGAAGCTCTGCCAGCGGCTTGTGAGGCGCACTTTTCTGCGCTC 1516  
 QY 1195 TGGGCTTGTGAGCTCTCTCTTCCCATGCTCAGGAGCACTGAGCTGAGCCCTACTT 1254





Db	666	CTGGTACCTGACCGGGCATTTGTAGAGCTGGAGGAGATGCTGTGGGGGCCGCAACAGGCTGG	725
Qy	1016	TGTGTACACTCTGCGCTCCAGCTATAGCTCTCTGTGATCCAAAGCAAGGTGACAGAACTTCA	1075
Db	726	TGTGTACACTCTGCGCTCCAGCTATAGCTCTCTGTGATCCAAAGCAAGGTGACAGAACTTCA	785
Qy	1076	GCTCTGTGTGGNGCCCAAAACCCAGAGATCTCCAGCCGACGACAGCAACCTCTGTGGACGCCA	1135
Db	786	GCTCTGTGTGGTGTGCCCCCAAAACCCAGAGATCTCCAGCCGACGACAGCAACCTCTGTGGACGCCA	845
Qy	1136	CCTGGCCCTTCAGCTCTGCCCCCAGCCGACGAGGCTT	1168
Db	846	CCTGGCCCTTCAGCTCTGCCCCCAGCCGACGAGGCTT	878
RESULT 15			
ID	ABN85394	standard; DNA; 882 BP.	
XX	ABN85394;		
XX	21-OCT-2002	(first entry)	
DE	Partial Human NOV14a DNA sequence.	162662711.	
KM	Human; NOV14a; cytosolic; Cardiant; Antiinflammatory; Immunosuppressive;		
KM	Antiallergic; Haemostatic; Anti-HIV; Antidiabetic; Anorectic;		
KM	Antiaesthetic; Nephrotropic; Hepatotropic; Neuroprotective; Nootropic;		
KM	Antibacterial; Virocidal; Antiparasitic; Relaxant; Anticonvulsant;		
KM	Gene Therapy; NOV; cancer; heart disease; inflammation;		
KM	autoimmune disorder; allergy; blood disorder; AIDS; diabetes;		
KM	obesity; asthma; IGA nephropathy; cirrhosis; arthritis; 162662711;		
KM	Alzheimer's disease; infection; stroke; muscular dystrophy; epilepsy;		
KM	wasting disorder; prostatic-like protein; chromosome 16; ds.		
OS	Homo sapiens.		
XX	WO20025704-A2.		
XX	18-JUL-2002.		
PF	09-JAN-2002; 2002WO-US00554.		
XX	09-JAN-2001; 2001US-260417P.		
PR	10-JAN-2001; 2001US-260831P.		
PR	28-FEB-2001; 2001US-272338P.		
PR	09-MAR-2001; 2001US-274876P.		
PR	16-APR-2001; 2001US-284704P.		
XX	(CURA-) CURAGEN CORP.		
PI	Padigaru M, Li L, Zehrhusen BD, Casman SJ, Shenoy S, Spytek KA;		
PI	Zhong G, Gangoli EA, Burgess CE, Patturajan M, Verner CM;		
PI	Taylor S, Tchenev VI, Miller CE, Guo X, Boldo FL, Grose KM;		
PI	Alsobrook JP, Gerlach V, Edinger S, Rothenberg ME, Ellerman K;		
PI	MacDougall J, Malyanar U, Millet I, Peyman J, Smithson G;		
PI	Gunther E, Stone DJ;		
DR	WPI; 2002-590674/63.		
CC	NOV14a is a proctasin-like protein, and the NOV14a coding sequences		
CC	localises to chromosome 16. The NOV proteins and NOV coding sequences of the		
CC	invention are useful for treating or preventing NOV-associated disorders		
CC	or in the manufacture of a medicament for treating the disorders, such as		
CC	cancer, heart disease, inflammation, autoimmune disorders, allergies,		
CC	pharmacogenomics -		
PT	Alzheimer's disease, and in chromosome mapping, tissue typing or		
PT	pharmacogenomics -		
PT	claim 9, Page 100; 358pp; English.		

CC blood disorders, AIDS, diabetes, obesity, asthma, IGA nephropathy,  
CC cirrhosis, arthritis, Alzheimer's disease, infections (e.g. bacterial  
CC viral, parasitic), stroke, muscular dystrophy, epilepsy, and other  
CC wasting disorders associated with chronic diseases.  
XX  
S0 Sequence 882 BP; 165 A; 304 C; 254 G; 155 T; 0 other;

Query Match	51.4%	Score 857.4	DB 24	length 882
Best Local Similarity	59.5%	Pred. No. 7.4e-175		
Matches 869; Conservative	1;	Mismatches 2;	Indels 1;	Gaps 1

296 GGGGAAGGGGCGAAGCTTCTCTGGGTGTGGCCCCCAAGCAGCATCAGTAGTGGCAGCC

356 AGTGCAGTGCACCGGTCAGTGCACCTGGCAGGTACGATCAACCTAGAAAGCCTCATGTG

67 AGTGCAGTCGCCGTCAGTGGCCCTGGCAGTCACATCACTTATGAAAGCGTCACATGTG

416 TGTGTGGCTCTCTGTTGTGAGCAGTGGGTGCTGTACGCTCTCACTTCTTCCAGGC

[illegible]

Db 187 GAGCACCAAGAGCCTATGAGGCTAAGCTGGGGGCCACACGCTAGACTCCCTACTCC

536 GAGGAGCGCAAGTTCAGCACCTGAGGACATCATCCCCCACCAGTACTCCAGAG

24 / GAGGACGCCAAGGTCAGCACCCTGAGGCAATCAATCCCCCAACCCAGCACTCCAGGAG

Db 307 GGCTCCCAAGGGCCGACATTGCACTCCTCCAACTCAGCAGACCCATCACTCTCCCGTAC

656 ATCCGGCCATCTGCTCCCTCCGAGCCAAGCCTCTTCCCAACGAGCTCCACTG

Db 367 ATCCGGCCATCTGCTCCCTCCGAGCCAAGCCTCTTCCCAACGGCTCCACTGACT

427 GTCACTGCTGGGCTCATGTGSCCCCCCTCATGTGACCTCTCGACGCGCCACGACCTGCAG

776 CAACTCGAGGTGCCTTGATCAGTCGTGAGACGTGGTAACTGCTGTATCAACATCGAGCGC

Dd 487 CAACTCGAGGTGCCTCTGATCAGTCGTGAGACGT-GTAACTGCCTGTACAAACATCGACGC

546 CAAGCCGAGGAGCCGCACTTTGTCCAGAGGATCATCTGCTGCTGCTGCTATCTGAGGG

896 GGGCAAGGACGCTTGCAGGCTGACTTGGGGCCACTCTTGCCTGTGAGGGTCT

Dp 606 GGGCAAGGACGCTGCGCAGGCTGACTCTGGGGGCCACTCTCTGCGCCCTGTGAGGGTCT

665 CTTGGCTTACGAGCGGCGAATGAGCCTGGGGAGAGTCCTGAGTGCCCTGGCAACAGCCCTGG  
938

1016 TGTGTAACCTCTGGCCTCCAGCTATGCTCTCGATCCAAAGCAAGTGCACAGAACTCCA

Db 726 TGTGTACTCTGGCCTCCAGCTATGCTCCTGATCCAAAGCAAGGTGACAGAACTCCA

706 GCGCTGATGGTGTCCCAACCCAGGAGTCCCAGCCCCGACGACAACCTGTGGCAGCCA  
10/6

1136 CCTGGCCTTCAgCTCTGCCCCAGCCCCAGGGCTT 1168

Db 846 CCTGGCCTTCACTCTGCCCAAGCCAGGGCGT 878

Search completed: December 15, 2003, 20:44:21

Job time : 487 secs

---

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using SW model

Run on: December 15, 2003, 20:44:27 ; Search time 552 Seconds  
(without alignments)  
10043.050 Million cell updates/sec

Title: US-09-925-301-208

Perfect score: 1668  
1 caccatgctcgttcggata.....tttggaaagttgaggagaa 1658

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2201672 seqs, 1661799599 residues

Total number of hits satisfying chosen parameters: 4403344

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-Processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications NA.\*

1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq.\*  
2: /cgn2\_6/ptodata/1/pubpna/PCT\_NEW\_PUB.seq.\*  
3: /cgn2\_6/ptodata/1/pubpna/US06\_NEW\_PUB.seq.\*  
4: /cgn2\_6/ptodata/1/pubpna/US06\_PUBCOMB.seq.\*  
5: /cgn2\_6/ptodata/1/pubpna/US07\_NEW\_PUB.seq.\*  
6: /cgn2\_6/ptodata/1/pubpna/PCTUS\_PUBCOMB.seq.\*  
7: /cgn2\_6/ptodata/1/pubpna/US08\_NEW\_PUB.seq.\*  
8: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq.\*  
9: /cgn2\_6/ptodata/1/pubpna/US09\_PUBCOMB.seq.\*  
10: /cgn2\_6/ptodata/1/pubpna/US09C\_PUBCOMB.seq.\*  
11: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq.\*  
12: /cgn2\_6/ptodata/1/pubpna/US09\_PUBCOMB.seq.\*  
13: /cgn2\_6/ptodata/1/pubpna/US10A\_PUBCOMB.seq.\*  
14: /cgn2\_6/ptodata/1/pubpna/US10B\_PUBCOMB.seq.\*  
15: /cgn2\_6/ptodata/1/pubpna/US10C\_PUBCOMB.seq.\*  
16: /cgn2\_6/ptodata/1/pubpna/US10\_NEW\_PUB.seq.\*  
17: /cgn2\_6/ptodata/1/pubpna/US60\_NEW\_PUB.seq.\*  
18: /cgn2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1662	99.6	1668	9	US-09-925-301-208
2	1549.4	92.9	3382	13	US-10-101-510-447
3	1537.4	92.2	1834	10	US-09-948-094-1
4	1537.4	92.2	1834	10	US-09-880-107-2214
5	1537.4	92.2	1834	10	US-09-967-768A-141
6	1537.4	92.2	1834	15	US-10-097-340-261
7	1420.8	85.2	1733	13	US-10-176-847-85
8	744.2	44.6	1797	15	US-10-109-616-1
9	594.2	35.6	596	9	US-09-922-217-931
10	594.2	35.6	596	10	US-09-833-263-931
11	594.2	35.6	596	14	US-10-025-380-931
12	487.6	29.2	521	15	US-10-066-543-1166
13	300	18.0	532	13	US-10-029-386-10805
14	294.6	17.7	543	13	US-10-029-386-7606
15	293.6	17.6	513	13	US-10-029-386-21306

US-09-925-301-208	US-10-029-386-24505	Sequence 24505, A				
16	272	16.3	273	13	US-10-041-400A-1	Sequence 1, Appl1
17	207.2	12.4	1613	14	US-10-041-264A-1	Sequence 1, Appl1
18	207.2	12.4	1613	14	US-10-041-264A-1	Sequence 1, Appl1
19	207.2	12.4	1613	14	US-10-041-264A-1	Sequence 1, Appl1
20	205.8	12.3	1130	14	US-10-041-400A-8	Sequence 8, Appl1
21	205.8	12.3	1130	14	US-10-041-264A-8	Sequence 8, Appl1
22	205.8	12.3	1130	14	US-10-041-091A-8	Sequence 8, Appl1
23	195.8	11.7	1130	14	US-10-041-655-8	Sequence 8, Appl1
24	195.8	11.7	1130	14	US-10-041-006A-8	Sequence 8, Appl1
25	192.8	11.6	780	15	US-10-266-035-1	Sequence 3, Appl1
26	192.8	11.6	980	15	US-10-266-035-1	Sequence 1, Appl1
27	192.8	11.6	980	15	US-10-040-647-30	Sequence 30, Appl1
28	192.8	11.6	1110	14	US-10-040-655-1	Sequence 1, Appl1
29	192.8	11.6	1110	14	US-10-041-006A-1	Sequence 1, Appl1
30	192.8	11.6	1129	13	US-10-137-870-221	Sequence 221, App
31	192.8	11.6	1129	13	US-10-140-018-221	Sequence 221, App
32	192.8	11.6	1129	13	US-10-140-021-221	Sequence 221, App
33	192.8	11.6	1129	13	US-10-140-274-221	Sequence 221, App
34	192.8	11.6	1129	13	US-10-140-471-221	Sequence 221, App
35	192.8	11.6	1129	13	US-10-140-807-221	Sequence 221, App
36	192.8	11.6	1129	13	US-10-140-924-221	Sequence 221, App
37	192.8	11.6	1129	13	US-10-140-926-221	Sequence 221, App
38	192.8	11.6	1129	13	US-10-141-658-221	Sequence 221, App
39	192.8	11.6	1129	13	US-10-141-702-221	Sequence 221, App
40	192.8	11.6	1129	13	US-10-141-704-221	Sequence 221, App
41	192.8	11.6	1129	13	US-10-142-421-221	Sequence 221, App
42	192.8	11.6	1129	13	US-10-142-421-221	Sequence 221, App
43	192.8	11.6	1129	13	US-10-142-421-221	Sequence 221, App
44	192.8	11.6	1129	13	US-10-142-421-221	Sequence 221, App
45	192.8	11.6	1129	13	US-10-143-033-221	Sequence 221, App

## ALIGNMENTS

RESULT 1  
US-09-925-301-208  
Sequence 208, Application US/09925301  
Patent No. US20020052308A1  
GENERAL INFORMATION:  
APPLICANT: Rosen et al.  
TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies  
FILE REFERENCE: PA106  
CURRENT APPLICATION NUMBER: US/09/925,301  
CURRENT FILING DATE: 2001-08-10  
PRIOR APPLICATION NUMBER: PCT/US00/05882  
PRIOR FILING DATE: 2000-03-08  
PRIOR APPLICATION NUMBER: 60/124,270  
PRIOR FILING DATE: 1999-03-12  
NUMBER OF SEQ ID NOS: 1694  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 208  
LENGTH: 1668  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (1505)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: misc feature  
LOCATION: (1565)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: misc feature  
LOCATION: (1598)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: misc feature  
LOCATION: (1620)  
OTHER INFORMATION: n equals a,t,g, or c  
US-09-925-301-208  
Query Match 99.6%; Score 1662; DB 9; Length 1668;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1668; Conservative 0; Mismatches 0; Gaps 0;

```

QY 1 CACACTGCTCGCTTGGAGATCTCCAGAGCGTCTCCGTTGGGCGCGCTCCCTGCTTAAAG 60
Db 1 CACACTGCTCGCTTGGAGATCTCCAGAGCGTCTCCGTTGGGCGCGCTCCCTGCTTAAAG 60
QY 61 GCCAGCCTTGGACACTTGTGCTGCTTCCAGCCGCGATTCTGGATTCCTTCCCTCTGAG 120
Db 61 GCCAGCCTTGGACACTTGTGCTGCTTCCAGCCGCGATTCTGGATTCCTTCCCTCTGAG 120
QY 121 CCACACTCTGGGCTCTGCTTGGACACCAACCGGCTTCCACTGGTGGTCCCTGGAG 180
Db 121 CCACACTCTGGGCTCTGCTTGGACACCAACCGGCTTCCACTGGTGGTCCCTGGAG 180
QY 181 TCTGCCCCAGAGGCGCTTGTCTGCGGCATGCGCMAAGAGGGGCTCTGGGCGCTGGAG 240
Db 181 TCTGCCCCAGAGGCGCTTGTCTGCGGCATGCGCMAAGAGGGGCTCTGGGCGCTGGAG 240
QY 241 GCTGGGCGGCTGTGGCAATCTGCTCTATCTTGGATTACTCGGTGGGAGCAGAGCGGA 300
Db 241 GCTGGGCGGCTGTGGCAATCTGCTCTATCTTGGATTACTCGGTGGGAGCAGAGCGGA 300
QY 301 AGGGGAGAAAGCTTCTGCGGCTGTGGCCCCCAGACGATCAAGAGTGGCAGCACTGC 360
Db 301 AGGGGAGAAAGCTTCTGCGGCTGTGGCCCCCAGACGATCAAGAGTGGCAGCACTGC 360
QY 361 AGTGGCGGCTGAGTGGGCGCTGGAGGTCAGATCACTATGAAGGCGTCAATGTGTGG 420
Db 361 AGTGGCGGCTGAGTGGGCGCTGGAGGTCAGATCACTATGAAGGCGTCAATGTGTGG 420
QY 421 TGGCTCTCTGCTGTGAGAGAGTGGGTGCTGTCACTGCTCTCTCCCAAGAGCA 480
Db 421 TGGCTCTCTGCTGTGAGAGAGTGGGTGCTGTCACTGCTCTCTCCCAAGAGCA 480
QY 481 CCACAAAGAGAGCTTATGAGGTCAAGCTGGGGGCGCAACAGCTAGACTCTCACTCGAGA 540
Db 481 CCACAAAGAGAGCTTATGAGGTCAAGCTGGGGGCGCAACAGCTAGACTCTCACTCGAGA 540
QY 541 CGCAGAGGTGAGACCTTGAAGAGCATCATCCCAAGAGCTGCTGAGAGAGGCTC 600
Db 541 CGCAGAGGTGAGACCTTGAAGAGCATCATCCCAAGAGCTGCTGAGAGAGGCTC 600
QY 601 CCAGGGGAGCATTCCTCTCCAACTCAGACAGACCCATCACTTCCGCGTACATCCG 660
Db 601 CCAGGGGAGCATTCCTCTCCAACTCAGACAGACCCATCACTTCCGCGTACATCCG 660
QY 661 GCCCATCTGCTCTCCCTGACAGCAACGCTCTTCCCAAGGCTCACTGCACTGTAC 720
Db 661 GCCCATCTGCTCTCCCTGACAGCAACGCTCTTCCCAAGGCTCACTGCACTGTAC 720
QY 721 TGGCTGGGCTCATGTGGGCGCTCTGAGAGCTCTGACGCGCAAGCACTGACAGCACT 780
Db 721 TGGCTGGGCTCATGTGGGCGCTCTGAGAGCTCTGACGCGCAAGCACTGACAGCACT 780
QY 781 CGAGGTGCTCTGATCAGTGTGAGAGCTGTAACTGCTGTACACATCGAGCGCAAGC 840
Db 781 CGAGGTGCTCTGATCAGTGTGAGAGCTGTAACTGCTGTACACATCGAGCGCAAGC 840
QY 841 CTGAGAGCGCGCACTTGTCCAGAGAGCATGTGTGTGCTGCTATGTGAGAGGCGCA 900
Db 841 CTGAGAGCGCGCACTTGTCCAGAGAGCATGTGTGTGCTGCTATGTGAGAGGCGCA 900
QY 901 AGGAGCGCTGCGCAGGGGAGCTGCTGGGGGCGCACTCTGCGCGCTGAGAGGCTCTGAG 960
Db 901 AGGAGCGCTGCGCAGGGGAGCTGCTGGGGGCGCACTCTGCGCGCTGAGAGGCTCTGAG 960
QY 961 ACCTGACGGGCAATTTGAGCTGGGAGATGCTGTGGGCGCGCAACAGGCTGTGTGT 1020
Db 961 ACCTGACGGGCAATTTGAGCTGGGAGATGCTGTGGGCGCGCAACAGGCTGTGTGT 1020
QY 1021 ACACTCTGGGCTCCAGCTATGCTCTCTGATCCAAAGCAAGGTACAGAACTCCAGCTTC 1080
Db 1021 ACACTCTGGGCTCCAGCTATGCTCTCTGATCCAAAGCAAGGTACAGAACTCCAGCTTC 1080

```

```

QY 1081 GTGTGGTGGCCCAAAACCAAGAGTCCAGCCCGAGACGAACTCTGTGGACCACTGG 1140
Db 1081 GTGTGGTGGCCCAAAACCAAGAGTCCAGCCCGAGACGAACTCTGTGGACCACTGG 1140
QY 1141 CCTTCAAGCTCTGCCCCAGGCGCTGTGAGGCCATCTTTTCCGCTCTGGAGCC 1200
Db 1141 CCTTCAAGCTCTGCCCCAGGCGCTGTGAGGCCATCTTTTCCGCTCTGGAGCC 1200
QY 1201 TGGCTCTGGGCTCTCTCTCCCATGAGGAGCACTGAGCTGAGCCCTTACCTTCAGAGA 1260
Db 1201 TGGCTCTGGGCTCTCTCTCCCATGAGGAGCACTGAGCTGAGCCCTTACCTTCAGAGA 1260
QY 1261 TGGATGATCACACTCAAGAGCAGAGGCTGTCTCTCCCTGATGAGCTTGGAGCCAGG 1320
Db 1261 TGGATGATCACACTCAAGAGCAGAGGCTGTCTCTCCCTGATGAGCTTGGAGCCAGG 1320
QY 1321 GCTGACTTGAACCACTCTTCTTCAAGACTCTGCGGAGAGCTGGGGCCCATCTTGAT 1380
Db 1321 GCTGACTTGAACCACTCTTCTTCAAGACTCTGCGGAGAGCTGGGGCCCATCTTGAT 1380
QY 1381 CTTTGAAGCCATTTCTTGGGTGTGCTTTTGGAGACATCACTGAGAGTCAGAGATTTA 1440
Db 1381 CTTTGAAGCCATTTCTTGGGTGTGCTTTTGGAGACATCACTGAGAGTCAGAGATTTA 1440
QY 1441 CTGCTGTAGCAATGCGCAGAGCTCTGCGCCCTCACCACCATGAGCCAGCCCATTGGG 1500
Db 1441 CTGCTGTAGCAATGCGCAGAGCTCTGCGCCCTCACCACCATGAGCCAGCCCATTGGG 1500
QY 1501 CGAGATCTTGGGAGTCTCTGGGACCTTGTGTATGAAAAGAACCTTGGGTTCCACCTGT 1560
Db 1501 CGAGATCTTGGGAGTCTCTGGGACCTTGTGTATGAAAAGAACCTTGGGTTCCACCTGT 1560
QY 1561 TTCTNAGAACCTGCTTCCGCGCGCGCTTCCAGACTGNATGAGCACTTTTTCGCGN 1620
Db 1561 TTCTNAGAACCTGCTTCCGCGCGCGCTTCCAGACTGNATGAGCACTTTTTCGCGN 1620
QY 1621 TTTCCCTGTGTTTGGGTGGGCAACTTTTGGAAATTGAGAGAA 1668
Db 1621 TTTCCCTGTGTTTGGGTGGGCAACTTTTGGAAATTGAGAGAA 1668

```

```

RESULT 2
US-10-101-510-447
; Sequence 447, Application US/10101510
; Publication No. US20030148295A1
; GENERAL INFORMATION:
; APPLICANT: WAN, JACKSON
; APPLICANT: WANG, YIXIN
; TITLE OF INVENTION: EXPRESSION PROFILES AND METHODS OF USE
; FILE REFERENCE: 15117.0012
; CURRENT APPLICATION NUMBER: US/10/101,510
; PRIOR FILING DATE: 2002-03-20
; PRIOR APPLICATION NUMBER: 60/276,947
; NUMBER OF SEQ ID NOS: 805
; SOFTWARE: Patent Ver. 2.1
; SEQ ID NO 447
; LENGTH: 3382
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-101-510-447

```

```

Query Match 92.9%; Score 1549.4; DB 13; Length 3382;
Best Local Similarity 97.8%; Pred. No. 0;
Matches 1626; Conservative 5; Mismatches 24; Indels 7; Gaps 6;
QY 1 CACACTGCTCGCTTGGAGATCTCCAGAGCTCTCCGTTGGGCGCGCTCCCTGCTTAAAG 60
Db 22 CACACTGCTCGCTTGGAGATCTCCAGAGCTCTCCGTTGGGCGCGCTCCCTGCTTAAAG 81
QY 61 GCCAGCCTTGGACACTTGTGCTGCTTCCAGCCGCGATTCTGGATTCCTTCCCTCTGAG 120
Db 82 GCCAGCCTTGGACACTTGTGCTGCTTCCAGCCGCGATTCTGGATTCCTTCCCTCTGAG 141

```



```

Db      202 TCGGCCCAAGGGGCGCTTGTCTT-AGGTAAGCCCAAGAGGGGCTCTGGGGCTGGGCA 260
Qy      241 GCTGGGGGCTGGGGCACTTCTGCTATCTTGGATTACTCGGTCCGGGACAGAGCCGA 300
Db      261 GCTGGGGGCTGGGGCACTTCTGCTATCTTGGATTACTCGGTCCGGGACAGAGCCGA 320
Qy      301 AGGGGCAAGAGCTCTGCGGGTGGGGCCCCCAAGCAGCATCAAGTGGGACAGTGC 360
Db      321 AGGGGCAAGAGCTCTGCGGGTGGGGCCCCCAAGCAGCATCAAGTGGGACAGTGC 380
Qy      361 AGTGGCGGCTAGTGGGCTTGGGAGGTGAGCATCACTATGAAGGCGTCATGTGTGG 420
Db      381 AGTGGCGGCTAGTGGGCTTGGGAGGTGAGCATCACTATGAAGGCGTCATGTGTGG 440
Qy      421 TGGCTCTCTGTGTCTGAGAGAGTGGGTCTGTCACTGCTCACTGCTTCCCGCAGCA 480
Db      441 TGGCTCTCTGTGTCTGAGAGAGTGGGTCTGTCACTGCTCACTGCTTCCCGCAGCA 500
Qy      481 CCAAGAGAGCTATGAGGTCAAGCTGGGGGGCCCAAGCTAGACTCTACTCCAGGA 540
Db      501 CCAAGAGAGCTATGAGGTCAAGCTGGGGGGCCCAAGCTAGACTCTACTCCAGGA 560
Qy      541 CGCCAAAGTGAAGACCTTGAAGACATATCCCCCAAGCTACCTCCAGAGGGCTC 600
Db      561 CGCCAAAGTGAAGACCTTGAAGACATATCCCCCAAGCTACCTCCAGAGGGCTC 620
Qy      601 CCAAGGCGCATGATGATCTCTCCAGCTCAGACAGCCATCACTTCTCCGCTATCCG 660
Db      621 CCAAGGCGCATGATGATCTCTCCAGCTCAGACAGCCATCACTTCTCCGCTATCCG 680
Qy      661 GCCCATCTGCTCTCTGCTGAGCAAGCGCTCTTCCCAAGCGGCTCCAGTCACTGTCA 720
Db      681 GCCCATCTGCTCTCTGCTGAGCAAGCGCTCTTCCCAAGCGGCTCCAGTCACTGTCA 740
Qy      721 TGGCTGGGGTCAATGTGGGCTTCTGAGTGGCTTCCAGCGCCCAAGCTCAGCAACT 780
Db      741 TGGCTGGGGTCAATGTGGGCTTCTGAGTGGCTTCCAGCGCCCAAGCTCAGCAACT 800
Qy      781 CAGAGTCCCTCTGATCACTGTGTGAGAGTGGTAACTGCTGTACAACTAGACCCCAAG 840
Db      801 CAGAGTCCCTCTGATCACTGTGTGAGAGTGGTAACTGCTGTACAACTAGACCCCAAG 860
Qy      841 CTGAGAGCGCGCATTTTGTCCAAAGAGACATGTGTGTGTGTGTGTGTGTGTGTGTGT 900
Db      860 CTGAGAGCGCGCATTTTGTCCAAAGAGACATGTGTGTGTGTGTGTGTGTGTGTGTGT 920
Qy      901 AGGACGCTTGCAGAGGTGATCTTGGGGGCCCACTTCTGCTGTGTGTGTGTGTGTGT 960
Db      920 AGGACGCTTGCAGAGGTGATCTTGGGGGCCCACTTCTGCTGTGTGTGTGTGTGTGT 979
Qy      961 ACCTGAAGGCGCATTTGTGAGTGGGAGATGTCTTGGGGGCCCAAGAGCTGTGTGTGT 1020
Db      980 ACCTGAAGGCGCATTTGTGAGTGGGAGATGTCTTGGGGGCCCAAGAGCTGTGTGTGT 1039
Qy      1021 ACATCTGAGCTTGCAGCTATGCTCTCTGATCAAGAGAGTGAAGAGTGTGTGTGTGTGT 1080
Db      1040 ACATCTGAGCTTGCAGCTATGCTCTCTGATCAAGAGAGTGAAGAGTGTGTGTGTGT 1099
Qy      1081 GTGTGTGTGCCCCAAACCCAGAGGTGCCAGGCCGACAGCAACTGTGTGTGTGTGTGT 1140
Db      1100 GTGTGTGTGCCCCAAACCCAGAGGTGCCAGGCCGACAGCAACTGTGTGTGTGTGTGT 1159
Qy      1141 CTTTCAAGCTCTGCCCCAGAGCCAGAGGCTTGTGAGAGCCCATCTTGTGTGTGTGTGTGT 1200
Db      1160 CTTTCAAGCTCTGCCCCAGAGCCAGAGGCTTGTGAGAGCCCATCTTGTGTGTGTGTGTGT 1219
Qy      1201 TGGCTTGGGGCTCTCTGCTCCCATAGGCTCAGAGAGCATGAGCTGGGCTCTTCCAGGA 1260
Db      1220 TGGCTTGGGGCTCTCTGCTCCCATAGGCTCAGAGAGCATGAGCTGGGCTCTTCCAGGA 1279
Qy      1261 TGAATGATCACTCAAGAGAGAGAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1320

```

```

Db      1280 TGAATGATCACTCAAGAGAGAGAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1339
Qy      1321 GCTGATCTTGAAGCACTCTCTCTTCCAGAGCTCTGCGGAGGCTGGGGCCCATCTTGAT 1380
Db      1340 GCTGATCTTGAAGCACTCTCTCTTCCAGAGCTCTGCGGAGGCTGGGGCCCATCTTGAT 1399
Qy      1381 CTTTGAAGCCATTTCTTGGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1440
Db      1400 CTTTGAAGCCATTTCTTGGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1459
Qy      1441 CTGCTGTGAGATAGGCGGAGAGCTCTGCGCCCTGAMCAGCATGAGAGCCCATTTGGS 1500
Db      1460 CTGCTGTGAGATAGGCGGAGAGCTCTGCGCCCTGAMCAGCATGAGAGCCCATTTGGS 1519
Qy      1501 CGAGTCTTGGGGAG--TCTGTGAGACTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1558
Db      1520 CGAGTCTTGGGGAGCTCTGCGAGCTCTGCGGAGCTGTGTGTGTGTGTGTGTGTGTGT 1578
Qy      1559 GTTTCCTGGAAGA CTGCTTCCGCGCCGCTTCCAGACTGATGAGACATTTTGTGTGT 1618
Db      1579 GTTTCCTGGAAGA CTGCTTCCGCGCCGCTTCCAGACTGATGAGACAT-TCTCTGTGC 1635
Qy      1619 CATTTCCTGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1660
Db      1636 CTTCTCTGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1677

RESULT 4
US-09-880-107-2214
US-09-880-107-2214, Application US/09880107
Sequence 2214, Application US/09880107
Patent No. US20020142981A1
GENERAL INFORMATION:
APPLICANT: Horne, Darci T.
APPLICANT: Vockley, Joseph G.
APPLICANT: Scherf, Dwe
APPLICANT: Gene Logic, Inc.
TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
FILE REFERENCE: 44921-5028-WO
CURRENT APPLICATION NUMBER: US/09/880,107
CURRENT FILING DATE: 2001-06-14
PRIOR APPLICATION NUMBER: US 60/211,379
PRIOR FILING DATE: 2000-06-14
PRIOR APPLICATION NUMBER: US 60/237,054
PRIOR FILING DATE: 2000-10-02
NUMBER OF SEQ ID NOS: 3950
SOFTWARE: Patentln Ver. 2.1
SEQ ID NO 2214
LENGTH: 1834
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: Genbank Accession No. US20020142981A1 L41351
US-09-880-107-2214

Query Match 92.2%; Score 1537.4; DB 10; Length 1834;
Best Local Similarity 97.8%; Pred. No. 0;
Matches 1625; Conservative 5; Mismatches 24; Indels 8; Gaps 7;

```

QY	241	GCTGGGGGCTGTGGCAATTCCTGCTTATCTTGGATTACTCCGGTCGGGAGCAGAGCGGA	300
Db	241	GCTGGGGGCTGTGGCAATTCCTGCTTATCTTGGATTACTCCGGTCGGGAGCAGAGCGGA	320
QY	301	AGGGGCAGAAGCTVCTCGGGGTGGGGCCCCCAACAGGACATCAAGGTGGAGCAATGCG	360
Db	321	AGGGGCAGAAGCTCTCTGGGGTGGGGCCCCCAACAGGACATCAAGGTGGAGCAATGCG	380
QY	361	AGTCGCCGATCAGTGGCCCTGGCAGGTGAGCATCACTATGAAAGCGTCCATGTGTCTGG	420
Db	381	AGTCGCCGATCAGTGGCCCTGGCAGGTGAGCATCACTATGAAAGCGTCCATGTGTCTGG	440
QY	421	TGGCTCTCTCGTGTCTGAGCAGTGGGTGCTGTCACTGTCTCACTGTCTTCCCGAGGACA	480
Db	441	TGGCTCTCTCGTGTCTGAGCAGTGGGTGCTGTCACTGTCTCACTGTCTTCCCGAGGACA	500
QY	481	CCACAAGGAAGCTATGAGGTCAACCTGGGGGCCCAACAGCTAGACTCTACTCCGAGGA	540
Db	501	CCACAAGGAAGCTATGAGGTCAACCTGGGGGCCCAACAGCTAGACTCTACTCCGAGGA	560
QY	541	CGCCAAAGTCAAGCAACCTGAAAGGAAATATCCCCACCCACAGCTACTCCAGAGGGCTC	600
Db	561	CGCCAAAGTCAAGCAACCTGAAAGGAAATATCCCCACCCACAGCTACTCCAGAGGGCTC	620
QY	601	CCAGGGGACATTGCACTCTTCCAACTGACAGAACCCATCACTTCTCCGCTACATTCG	660
Db	621	CCAGGGGACATTGCACTCTTCCAACTGACAGAACCCATCACTTCTCCGCTACATTCG	680
QY	661	GCCCATCTGCCCTCCCTGAGGCCAAGGCTCTTCTCCCAAGGCTCTCACTGACCTGTAC	720
Db	681	GCCCATCTGCCCTCCCTGAGGCCAAGGCTCTTCTCCCAAGGCTCTCACTGACCTGTAC	740
QY	721	TGGCTGGGGGTATGTGGGCCCCCTCAGTAGGCTCTCTGAAGCCCAAGCCACTGACAGCACT	780
Db	741	TGGCTGGGGGTATGTGGGCCCCCTCAGTAGGCTCTCTGAAGCCCAAGCCACTGACAGCACT	800
QY	781	CGAGGTGCTCTGTATCACTGCTGTGAGACGCTGTGAATCTGCTGTATCAACAATCGACCGCAAGC	840
Db	801	CGAGGTGCTCTGTATCACTGCTGTGAGACGCTGTGAATCTGCTGTATCAACAATCGACCGCAAGC	859
QY	841	CTTGAGGAGCCGCACTTTGTCCAAAGAGACATGTGTGTGTCTGTGGCTATGTGAGGGGGGACA	900
Db	860	CTTGAGGAGCCGCACTTTGTCCAAAGAGACATGTGTGTGTCTGTGGCTATGTGAGGGGGGACA	919
QY	901	AGGACGCTGTCAGGGGTGACTCTGGGGGGGCCACTCTGCTGCCCTGTGAGGGGTCTCTGT	960
Db	920	AGGACGCTGTCAGGGGTGACTCTGGGGGGGCCACTCTCTGCTGCCCTGTGAGGGGTCTCTGT	979
QY	961	ACTTGACGGGCAATTGTGAGCTGGGGAGATGTGCTGTGGGGCCCGCAACAGGCTGGTGTGT	1020
Db	980	ACTTGACGGGCAATTGTGAGCTGGGGAGATGTGCTGTGGGGCCCGCAACAGGCTGGTGTGT	1039
QY	1021	ACACTCTGGCCTCCAGCTATGCTCTCTTGGATCCAAAGCAAGGTGACAGAACTCCAGCCTC	1080
Db	1040	ACACTCTGGCCTCCAGCTATGCTCTCTTGGATCCAAAGCAAGGTGACAGAACTCCAGCCTC	1099
QY	1081	GTTGTGTGGCCCAAAACCCAGGAGTCCAGGCCCGACAGCAACTCTGTGTGGCAGCACTCG	1140
Db	1100	GTTGTGTGGCCCAAAACCCAGGAGTCCAGGCCCGACAGCAACTCTGTGTGGCAGCACTCG	1159
QY	1141	CTTTACAGCTCTGCCCCAAGCCAGGGCTTGTGAGGGCCACTCTTCTTCTGTGCTCTGGGCTC	1200
Db	1160	CTTTACAGCTCTGCCCCAAGCCAGGGCTTGTGAGGGCCACTCTTCTTCTGTGCTCTGGGCTC	1219
QY	1201	TGGCTCTGGGCTCTCTCTCCCAATGGCTCAAGCAGACCTGAGCTTGGCCCTTCCAGGA	1260
Db	1220	TGGCTCTGGGCTCTCTCTCCCAATGGCTCAAGCAGACCTGAGCTTGGCCCTTCCAGGA	1279
QY	1261	TGGATGCACTACACTCAAGGACAGAGAGCTGTCTTCCCTGATGTGGCTTTTGGACCAAG	1320
Db	1280	TGGATGCACTACACTCAAGGACAGAGAGCTGTCTTCCCTGATGTGGCTTTTGGACCAAG	1339

QY	1321	UCCCTGACTGAGCCACTCTCTCTTCAAGACTCTGCGGAGAGGCTGGGGCCCCACTTGTAT	1380
Db	1340	GCCGACTGTGAGCCACTCTCTCTTCAAGACTCTGCGGAGAGGCTGGGGCCCCACTTGTAT	1399
QY	1381	CTTTGAGCCCATCTTCTGAGGTGTCTTTTGGAGCATCATCTAGAGTCAAGAGTTTAA	1440
Db	1400	CTTTGAGCCCATCTTCTGAGGTGTCTTTTGGAGCATCATCTAGAGTCAAGAGTTTAA	1459
QY	1441	CTGCTGTAGCAATGAGGCAGAGCCTCTGAGCCCTCAGACCAATGAGACAGCCCATTTGS	1500
Db	1460	CTGCTGTAGCAATGAGGCAGAGCCTCTGAGCCCTCAGACCAATGAGACAGCCCATTTGS	1519
QY	1501	CGAGNCTCTGAGGAG--TCTGGGACCTTGATATGAAATGAGCCCTGGGTTTCCACT	1558
Db	1520	CGAGCTCCTGGGGAGACTCTGTGGAGCCTTGAGCTATGAAATGAGCCCT--GGCTCCCACT	1578
QY	1559	GTTTCTGAAAGACTGTCTTCCCGGCCCCGCTTCCAGACTGATGATAGACATTTTGTGTC	1618
Db	1579	GTTTCTGAAAGACTGC--TCCCGGCCCCGCTTCCAGACT--GATGAGACA--TCTCTGTGC	1635
QY	1619	CNTTTCCTGTGTTTGGGGTGGGAACTTTTGGAAAGTTT	1660
Db	1636	CTCTCTCCCTGTGTCTGGGCTGGGGCCACCTTTGTGTGAGCTT	1677

```

RESULT 5
US-09-967-768A-141
; Sequence 141, Application US/09967768A
; Patent No. US2002015087A1
; GENERAL INFORMATION:
; APPLICANT: Augustus, Meena
; TITLE OF INVENTION: Cancer Gene Determination and Therapeutic Screening Using Signature
; TITLE OF INVENTION: Sets
; FILE REFERENCE: 689290-72
; CURRENT APPLICATION NUMBER: US/09/967,768A
; CURRENT FILING DATE: 2001-09-28
; PRIOR APPLICATION NUMBER: US/60/236,109
; PRIOR FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: US/60/236,034
; PRIOR FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: US/60/236,111
; PRIOR FILING DATE: 2000-09-28
; NUMBER OF SEQ ID NOS: 325
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 141
; LENGTH: 1834
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-967-768A-141

Query Match      92.2%; Score 1537.4; DB 10; Length 1834;
Best Local Similarity 97.8%; Pred. No. 0;
Matches 1625; Conservative 5; Mismatches 24; Indels 8; Gaps 7;

QY.      1  CACACTGCTCGCTTGATGATCTCCAGAGCGTCTCCGGTCCGCGCCGCTCCCTGCTTAGAG 60
          |||
Db        22  CACACTGCTCGCTTGATGATCTCCAGAGCGTCTCCGGTCCGCGCCGCTCCCTGCTTAGAG 81

QY        61  GCCAGCCTTGGACACTCTGCGCCCTTTCAGCCCGCGATTTCTGGATCTTCCCTCTGAG 120
          |||
Db        82  GCCAGCCTTGGACACTCTGCGCCCTTTCAGCCCGCGATTTCTGGATCTTCCCTCTGAG 141

QY       121  CCAACATTTGGGTCTCGCTTGGACACACACCCCAAGGCTTCTACTCTTGCTGCTCGAG 180
          |||
Db       142  CCAACATTTGGGTCTCGCTTGGACACACACCCCAAGGCTTCTACTCTTGCTGCTCGAG 201

QY       181  TCTGCCCCAGGGGCGCTTGTCTCTGAGGCATGAGCCMAAGAGGGAGTCTTGAGGCTCGAGCA 240
          |||
Db       202  TCTGCCCCAGGGGCGCTTGTCTCT-GGCATGAGCCCAAGAGGGAGTCTTGAGGCTCGAGCA 260

QY       241  GCTGAGGAGCTGTGGCATTTCTGCTCTATCTTGGATTACTCGGAGTGGAGACAGAGCCGGA 300
          |||
Db       261  GCTGAGGAGCTGTGGCATTTCTGCTCTATCTTGGATTACTCGGAGTGGAGACAGAGCCGGA 320

```



QY 301 AGGGGAGAGAGCTTCCGAGGAGTGGGCCCCCAAGACGATCATAGGTGGAGAGTGC 360  
 Db 321 AGGGGAGAGAGCTTCCGAGGAGTGGGCCCCCAAGACGATCATAGGTGGAGAGTGC 380  
 QY 361 AGTGGCCGAGTGGGAGTGGGAGTGGGAGTGGGAGTGGGAGTGGGAGTGGGAGTGGG 420  
 Db 381 AGTGGCCGAGTGGGAGTGGGAGTGGGAGTGGGAGTGGGAGTGGGAGTGGGAGTGGG 440  
 QY 421 TGGCTCTCTGT 480  
 Db 441 TGGCTCTCTGT 500  
 QY 481 CCAGAGAGAGAGCTTATAGAGTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 540  
 Db 501 CCAGAGAGAGAGCTTATAGAGTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 560  
 QY 541 CGCCAGAGTGGAGAGCTTGAAGAGATCATCCCGACCCGAGCTTACTCCAGAGAGGCTC 600  
 Db 561 CGCCAGAGTGGAGAGCTTGAAGAGATCATCCCGACCCGAGCTTACTCCAGAGAGGCTC 620  
 QY 601 CCAGAGGAGAGATTTGACCTCTTCCAACTCAGAGAGCCATCACTTCTCCGCTTACATCCG 660  
 Db 621 CCAGAGGAGAGATTTGACCTCTTCCAACTCAGAGAGCCATCACTTCTCCGCTTACATCCG 680  
 QY 661 GCCCATCTGCTCCCTGAGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 720  
 Db 681 GCCCATCTGCTCCCTGAGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 740  
 QY 721 TGGCTGGGAGTCAATGTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 780  
 Db 741 TGGCTGGGAGTCAATGTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 800  
 QY 781 CGAGAGTCTCTGT 840  
 Db 801 CGAGAGTCTCTGT 860  
 QY 841 CTAGAGAGAGGAGCTTGT 900  
 Db 860 CTAGAGAGAGGAGCTTGT 920  
 QY 901 AGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 960  
 Db 920 AGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 980  
 QY 961 ACCTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1020  
 Db 980 ACCTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1040  
 QY 1021 ACACTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1080  
 Db 1040 ACACTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1100  
 QY 1081 GT 1140  
 Db 1100 GT 1160  
 QY 1141 CTTTCACTCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1200  
 Db 1160 CTTTCACTCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1220  
 QY 1201 TGGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1260  
 Db 1220 TGGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1280  
 QY 1261 TGGATGATCATCACTCAAGAGCAGAGAGCTGTGTCTTCTCTGTGTGTGTGTGTGTGTGTGT 1320  
 Db 1280 TGGATGATCATCACTCAAGAGCAGAGAGCTGTGTCTTCTCTGTGTGTGTGTGTGTGTGTGT 1340  
 QY 1321 GCGTGAATTTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1380  
 Db 1340 GCGTGAATTTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1400

QY 1381 CTTTGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1440  
 Db 1400 CTTTGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1460  
 QY 1441 CTGCTGT 1500  
 Db 1460 CTGCTGT 1520  
 QY 1501 CGAG 1560  
 Db 1520 CGAG 1580  
 QY 1559 GTTCTTGAAGAGCTGT 1620  
 Db 1579 GTTCTTGAAGAGCTGT 1640  
 QY 1619 CTTTCTGT 1680  
 Db 1636 CTTTCTGT 1700  
 RESULT 6  
 US-10-097-340-261  
 ; Sequence 261, Application US/10097340  
 ; Publication No. US20030087250A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: John MONAHAN  
 ; APPLICANT: Manjula GANNANARAPU  
 ; APPLICANT: Sebastian HOERCH  
 ; APPLICANT: Shubhangi KAMATKAR  
 ; APPLICANT: Steve G. KOVATS  
 ; APPLICANT: Rachel E. MEYERS  
 ; APPLICANT: Michael MORRISSEY  
 ; APPLICANT: Peter OLANDT  
 ; APPLICANT: Ami SEN  
 ; APPLICANT: Peter VEIBY  
 ; APPLICANT: Gordon B. MILLS  
 ; APPLICANT: Robert C. BAST, Jr.  
 ; APPLICANT: Karen LU  
 ; APPLICANT: Rosemarie SCHMANDT  
 ; APPLICANT: Xumel ZHAO  
 ; APPLICANT: Karen GLATT  
 ; TITLE OF INVENTION: Nucleic Acid Molecules and Proteins For The Identification,  
 ; TITLE OF INVENTION: Assessment, Prevention, and Therapy of Ovarian Cancer  
 ; FILE REFERENCE: MRI-030  
 ; CURRENT APPLICATION NUMBER: US/10/097,340  
 ; CURRENT FILING DATE: 2002-03-14  
 ; PRIOR APPLICATION NUMBER: 60/276,025  
 ; PRIOR FILING DATE: 2001-03-14  
 ; PRIOR APPLICATION NUMBER: 60/325,149  
 ; PRIOR FILING DATE: 2001-09-26  
 ; PRIOR APPLICATION NUMBER: 60/276,026  
 ; PRIOR FILING DATE: 2001-03-14  
 ; PRIOR APPLICATION NUMBER: 60/324,967  
 ; PRIOR FILING DATE: 2001/09/26  
 ; PRIOR APPLICATION NUMBER: 60/311,732  
 ; PRIOR FILING DATE: 2001-08-10  
 ; PRIOR APPLICATION NUMBER: 60/325,102  
 ; PRIOR FILING DATE: 2001-09-26  
 ; PRIOR APPLICATION NUMBER: 60/323,580  
 ; PRIOR FILING DATE: 2001-09-19  
 ; NUMBER OF SEQ ID NOS: 363  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 261  
 ; LENGTH: 1834  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 ; US-10-097-340-261  
 Query Match 92.2%; Score 1537.4; DB 15; Length 1834;  
 Best Local Similarity 97.8%; Pred. No. 0;  
 Matches 1625; Conservative 5; Mismatches 24; Indels 8; Gaps 7;



QY	1	CACACTGCTCCCTTGAGATATCCACAGGGGCTCCCGTTGCGGGCCGCTCCCTCCCTTAAAG	60
Db	22	CACACTGCTCCGTTGGATTACTCCAGGGGCTCTCCGTTGGGGCCGCTCCCTGCTTTAAG	81
QY	61	GCCAGCCTTGGACATTGCTGCCCCCTTTCAGCCCGGATTCGTGGATCCTTCCCTTGAAG	120
Db	82	GCCAGCCTTGGACACTTCTGTGCCCCCTTTCAGCCCGGATTCGTGGATCCTTCCCTTGAAG	141
QY	121	CCAACTATTGGGATCCTGCTTCCGATCCACCCCAAGGCTTCTACTTGGCTGCTGGAG	180
Db	142	CCAACTATTGGGATCCTGCTTCCGATCCACCCCAAGGCTTCTACTTGGCTGCTGGAG	201
QY	181	TCTGCCCCAGGGGACCCTTGTCTGTGGGACATGGCCAAAGAGGGGGCTGTGGGGCTGGGCA	240
Db	202	TCTGCCCCAGGGGACCCTTGTCTGT - GGCATGGCCAAAGAGGGGGCTGTGGGGCTGGGCA	260
QY	241	GCTGGGGGCTGTGGGCACTTGTCTTACTTGTGAATTACTCGGCTCGGGGACAGAGCGGA	300
Db	261	GCTGGGGGCTGTGGGCACTTGTCTTACTTGTGAATTACTCGGCTCGGGGACAGAGCGGA	320
QY	301	AGGGGCGAAGCTTCTGTGCGGCTGTGGCCCCCAAGCAGCATCACAGTGGCAGCATGTGC	360
Db	321	AGGGGCGAAGACTCCCTGTGCGGCTGTGGCCCCCAAGCAGCATCACAGTGGCAGCATGTGC	380
QY	361	AGTGGCGGCTGAGTGGGCCCTGGCAGGTGAGATATCACTATGAGGGGGCTCAATGTGTGG	420
Db	381	AGTGGCGGCTGAGTGGGCCCTGGCAGGTGAGATATCACTATGAGGGGGCTCAATGTGTGG	440
QY	421	TGGCTCTTCTGTTCTTGAGCAGTGGGCTGTGCAGCTGTCACTGTCTTCCGAGGAGCA	480
Db	441	TGGCTCTTCTGTTCTTGAGCAGTGGGCTGTGCAGCTGTCACTGTCTTCCGAGGAGCA	500
QY	481	CCACAGAGAGCCTATGAGTTCAACTGGGGGCCCAACAGCTAGACTCTTACTCCGAGGA	540
Db	501	CCACAGAGAGACCTATGAGTTCAACTGGGGGCCCAACAGCTAGACTCTTACTCCGAGGA	560
QY	541	CGCCAGGTCAGACACCCTGAAGAGCATATACCCCAACCCAGCTACCTCCAGAGAGGCTC	600
Db	561	CGCCAGGTCAGACACCCTGAAGAGCATATACCCCAACCCAGCTACCTCCAGAGAGGCTC	620
QY	601	CCAGGGCGACATTTGCACTCTTCCAACTCAGCAGAGCCCATTCACCTTCTCCCGTATACCG	660
Db	621	CCAGGGCGACATTTGCACTCTTCCAACTCAGCAGAGCCCATTCACCTTCTCCCGTATACCG	680
QY	661	GCCCATCTGCTCCCTGTGAGCGCAACGCGCTCTTCCCAAGCGGCTTCACTGCACTGTAC	720
Db	681	GCCCATCTGCTCCCTGTGAGCGCAACGCGCTCTTCCCAAGCGGCTTCACTGCACTGTAC	740
QY	721	TGGCTGGGGTATATGTGGCCCCCTTCAGTAGGCTCTTGAAGCCCAAGCCACTTGCAGACT	780
Db	741	TGGCTGGGGTATATGTGGCCCCCTTCAGTAGGCTCTTGAAGCCCAAGCCACTTGCAGACT	800
QY	781	CGAGGTGCTGTATGACGTGTTGAGCGTTGTAACCTGTGTGAACAATTCGACGCAAGC	840
Db	801	CGAGGTGCTGTATGACGTGTTGAGCGT - GTAACCTGTGTGAACAATTCGACGCAAGC	859
QY	841	CTGAGAGAGCCGACATTTGTTCGAAGAGCATATGTGTGTGTCTGTATGTGAGAGGGGCA	900
Db	860	CTGAGAGAGCCGACATTTGTTCGAAGAGCATATGTGTGTGTCTGTATGTGAGAGGGGCA	919
QY	901	AGAGAGCGCTTGCAGAGGTATCTTGGGGGCCCATCTTCTGTGCTGTGAGAGGCTCTTGT	960
Db	920	AGAGAGCGCTTGCAGAGGTATCTTGGGGGCCCATCTTCTGTGCTGTGAGAGGCTCTTGT	979
QY	961	ACCTGACGGGCAATTGTGAGCTGSGGAGATGTCTGTGGGGCCCGCAACAGGCTGGTGT	1020
Db	980	ACCTGACGGGCAATTGTGAGCTGSGGAGATGTCTGTGGGGCCCGCAACAGGCTGGTGT	1039
QY	1021	ACACTTGGCTTCCAGCTATGCTCTTGTGATCCAAAGCAAGGTGACAGAACTCCAGCCTC	1088
Db	1040	ACACTTGGCTTCCAGCTATGCTCTTGTGATCCAAAGCAAGGTGACAGAACTCCAGCCTC	1099
QY	1081	GTTGTGTGTGCCCCAACCAGAGTTCAGCCCGACAGCAACTCTGTGTGGACCACTTGG	1140

Db 1100 GTGTGGTGGCCCAACCCAGAGATCCAGGCCACAGCAACTCTGTGGAGACCACTGG 1159

QY 1141 CCTTCAGCTCTGCCCCAGGCCAGGGCTTGTGAGGCCCATCTTTTCTGCTCTGAGGCC 1200

Db 1160 CCTTCAGCTCTGCCCCAGGCCAGGGCTTGTGAGGCCCATCTTTTCTGCTCTGAGGCC 1219

QY 1201 TTGGCTCTGGGGCTCTCTCTCCCACTGGGTCAAGCAGACATGAGTGGGCCCTTACTTCCAGA 1260

Db 1220 TTGGCTCTGGGGCTCTCTCTCCCACTGGGTCAAGCAGACATGAGTGGGCCCTTACTTCCAGA 1279

QY 1261 TTGATGATACACACTCAAGGACAGAGCCCTGATCTTCCCTGATGGCCCTTTGAGCCAGG 1320

Db 1280 TTGATGATACACACTCAAGGACAGAGCCCTGATCTTCCCTGATGGCCCTTTGAGCCAGG 1339

QY 1321 GCTGATCTTGAGCCACTCTCTCTTCAAGACATCTGCGGAGGCTGAGGCCCATCTTGTAT 1380

Db 1340 GCTGATCTTGAGCCACTCTCTCTTCAAGACATCTGCGGAGGCTGAGGCCCATCTTGTAT 1399

QY 1381 CTTTGAAGCCATCTCTCTGGGTGTCTTTTGGAGCATTACTGAAGTCAAGAGTTTAA 1440

Db 1400 CTTTGAAGCCATCTCTCTGGGTGTCTTTTGGAGCATTACTGAAGTCAAGAGTTTAA 1459

QY 1441 CTGCGATGAGCAATGGACAGAGCCTCGGGCCCTGACACATGAGCCAGGCCATTGGG 1500

Db 1460 CTGCGATGAGCAATGGACAGAGCCTCGGGCCCTGACACATGAGCCAGGCCATTGGG 1519

QY 1501 CGAGNTCTGGGGAG--TCTCGGAGACTTGGTATGAAATGAGCCCTGGGTTCCACACT 1558

Db 1520 CGAGNTCTGGGGAGCCTCGGGACCCCTTGGCTATGAAATGAGCCCT--GGCTCCACACT 1578

QY 1559 GTTTCGNAAGACTGCTTCCCGGCCCGCTTCCCAAGCTNATGAGACATTTTGTGC 1618

Db 1579 GTTTCGNAAGACTGCT--TCCCGGCCCGCTTCCCAAGCT--GATGAGACACA--TCTCTGTGC 1635

QY 1619 CTTTCCCTGTGTTTTGGGTGGGCACTTTTGAAGTTT 1660

Db 1636 CTTTCCCTGTGTTTTGGGTGGGCACTTTTGAAGTTT 1677

RESULT 7

US-10-176-847-85

Sequence 85, Application US/10176847

Publication No. US20030068636A1

GENERAL INFORMATION:

APPLICANT: Velby, Peter Ole

TITLE OF INVENTION: COMPOSITIONS, KITS, AND METHODS FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY OF BREAST TITILE OF INVENTION: AND OVARIAN CANCER

FILE REFERENCE: MRI-039

CURRENT APPLICATION NUMBER: US/10/176,847

CURRENT FILING DATE: 2002-06-21

NUMBER OF SEQ. ID NOS: 112

SOFTWARE: PatsSEO for windows Version 4.0

SEQ ID NO 85

LENGTH: 1733

TYPE: DNA

ORGANISM: Homo sapiens

US-10-176-847-85

Query Match 85.2%; Score 1420.8; DB 15; Length 1733;

Best Local Similarity 97.6%; Pred. No. 0;

Matches 1498; Conservative 5; Mismatches 25; Indels 7; Gaps 6;

QY 128 CTGGGTCTGCGCTTGAGACCAACCCCAAGCTTCTCACTGTGGTCCCTGGAGTCTGGCC 187

Db 18 CTGGGTCTGCGCTTGAGACCAACCCCAAGCTTCTCACTGTGGTCCCTGGAGTCTGGCC 77

QY 188 CAGGGGCCCTTGTCTGAGGCAATGCGCMAAGAGGGGTCTGAGGCGCTGAGGAGCTGAGG 247

Db 78 CAGGGGCCCTTGTCTGAGGCAATGCGCMAAGAGGGGTCTGAGGCGCTGAGGAGCTGAGG 137

QY 248 GCTGTGGCAATCTGCTATCTTGATTACTCGGTGAGGACAGAGCGAAGAGGGGCA 307

Db	138	CTGTGGCCATTCTGCTCTATCTTGATTACTCCGGTCAAGGACAGAGCGGAAGGGGCA	197
Qy	308	GAAGCTTCTGCGGTGTGGCCCCCAAGCAGCATCAAGGTGGCAGCATGACGTGCC	367
Db	198	GAAGCTCCCTGCGGTGTGGCCCCCAAGCAGCATCAAGGTGGCAGCATGACGTGCC	257
Qy	368	GGTCAATGTGCCCCCTGGCAGGTAGCATATACCTATGAAGGCGCTCATGTGTGTGGTCT	427
Db	258	GGTCAATGTGCCCCCTGGCAGGTAGCATATACCTATGAAGGCGCTCATGTGTGTGGTCT	317
Qy	428	CTCGGTCTGAGCAGTGGGTGCTGTGACGTCTCACTACGTCTTCCCAAGCAGCACCAAG	487
Db	318	CTCGGTCTGAGCAGTGGGTGCTGTGACGTCTCACTACGTCTTCCCAAGCAGCACCAAG	377
Qy	488	GAAGCCTATGAGGTCAAGCTGGGGGCCCAAGCAGTGACTCTCTATCCGAGAGCGCAAG	547
Db	378	GAAGCCTATGAGGTCAAGCTGGGGGCCCAAGCAGTGACTCTCTATCCGAGAGCGCAAG	437
Qy	548	GTCAAGCACCCTGAAGAGACATATCCCCCAAGCTACCTCCAGAGAGGGTCCCAAGGC	607
Db	438	GTCAAGCACCCTGAAGAGACATATCCCCCAAGCTACCTCCAGAGAGGGTCCCAAGGC	497
Qy	608	GACATTTGACCTCTCCAACTCAGAGACCCATACCTTCTCCGGTACATCGGGCCATC	667
Db	498	GACATTTGACCTCTCCAACTCAGAGACCCATACCTTCTCCGGTACATCGGGCCATC	557
Qy	668	TGCGTCCCTGAGGCAACGCTCTCTTCCCAAGCGGCTTCCACTGACCTGTCACTGGCTGG	727
Db	558	TGCGTCCCTGAGGCAACGCTCTCTTCCCAAGCGGCTTCCACTGACCTGTCACTGGCTGG	617
Qy	728	GGTATATGTGCCCCCTCAGTGAAGCTCTCTGAAGGCCCAAGCCACTGACCAACTGAGGTG	787
Db	618	GGTATATGTGCCCCCTCAGTGAAGCTCTCTGAAGGCCCAAGCCACTGACCAACTGAGGTG	677
Qy	788	CCTCGAATCAGTCGTGAGACGTGGTAACTGACTGTGAACAATTCGACGCGCAAGCTGAGGA	847
Db	678	CCTCGAATCAGTCGTGAGACGTGGTAACTGACTGTGAACAATTCGACGCGCAAGCTGAGGA	736
Qy	848	GCCGCACTTTGTCCAAAGAGACATGATGTGTGCTGAGCTATGTGAGGGGGCAAGACGC	907
Db	737	GCCGCACTTTGTCCAAAGAGACATGATGTGTGCTGAGCTATGTGAGGGGGCAAGACGC	796
Qy	908	CTGCCAGGGTGACTCTGTGGGGGCCACTCTCCTGCCCTGTGAGGGGTCTCTGTATCTGAC	967
Db	797	CTGCCAGGGTGACTCTGTGGGGGCCACTCTCCTGCCCTGTGAGGGGTCTCTGTATCTGAC	856
Qy	968	GGGCAATTGTAGCTGTGGGGAGATGCTCTGTGGGGGCCCGCAACAGGCTGGTGTGTACCTCT	1027
Db	857	GGGCAATTGTAGCTGTGGGGAGATGCTCTGTGGGGGCCCGCAACAGGCTGGTGTGTACCTCT	916
Qy	1028	GGCCTCCAGCTATGCTCTCTGGATCCAAAGCAAGGTGACAGAACTCCAGCCTCGTGTGT	1087
Db	917	GGCCTCCAGCTATGCTCTCTGGATCCAAAGCAAGGTGACAGAACTCCAGCCTCGTGTGT	976
Qy	1088	GCCCCAAACCCAGAGATCCAGCCCGCAGCAGCAACTCTGTGGCAGCCACTGGGCTTCAAG	1147
Db	977	GCCCCAAACCCAGAGATCCAGCCCGCAGCAGCAACTCTGTGGCAGCCACTGGGCTTCAAG	1036
Qy	1148	CTTGCCCCCAGCCCAAGGGCTTGTGAGGCCCATCTTTTCTGCTCTGGGCTGTGGCTCT	1207
Db	1037	CTTGCCCCCAGCCCAAGGGCTTGTGAGGCCCATCTTTTCTGCTCTGGGCTGTGGCTCT	1096
Qy	1208	GGGCTCTCTTCCCATGAGTCAAGGAGACATGAGCTGGCCCTACTTCCAGAGATGATGC	1267
Db	1097	GGGCTCTCTTCCCATGAGTCAAGGAGACATGAGCTGGCCCTACTTCCAGAGATGATGC	1156
Qy	1268	ATCACTCTCAAGAGACAGAGGCTGTGCTTCTCCTGATAGGCTTTTGAACCAAGGCGCTGAC	1327
Db	1157	ATCACTCTCAAGAGACAGAGGCTGTGCTTCTCCTGATAGGCTTTTGAACCAAGGCGCTGAC	1216
Qy	1328	TTGAGCACTCTTCTTCAAGACTCTGCGGAGGCTGGGGCCCATCTTGAATCTTTGAG	1388

Db	1217	TTGAGCAGACGTCCTCTTTCAGAGACTCTGCGGAGAGCGCTGGGGGCCCCACATCTTGATCTTTGAG	1276
Qy	1388	CCCAATTTCTTGGGTGGTGGCTTTTGGGACCATATCATGAGAGTCAAGAGTTTACTGCGCTG	1447
Db	1277	CCAAATTTCTTGGGTGGTGGCTTTTGGGACCATATCATGAGAGTCAAGAGTTTACTGCGCTG	1336
Qy	1448	TAGCAATGGCCAGAGCCTCTGGGCCCTCAGCCACCATGAGCAGACGCCATTTGGSCGAGATC	1507
Db	1337	TAGCAATGGCCAGAGCCTCTGGGCCCTCAGCCACCATGAGCAGACGCCATTTGGSCGAGATC	1396
Qy	1508	CTGGGGAG--TCTGGGAGCCTTGGGYTATGAAGATGACCCCTGGGGTCCCACTGTTCTN	1565
Db	1397	CTGGGGAGCTCTGGGAGCCCTTGGCTATGAAGATGAGCCCT--GGCTCCACACTGTTCTG	1455
Qy	1566	GAGACACTGTTCCCGGCGCCGCGCTTCCAGCATNGATAGAGACATTTTTTGTGCNNTTCC	1625
Db	1456	GAAACACTGC--TCCGGCGCCGCGCTTCCAGCAT--GATAGAGACA--TCTCTCGCCCTCTCC	1512
Qy	1626	CTGNGTTTGGGTGGGCAACTTTTGGAAAGTTT	1660
Db	1513	CTGNGTTTGGGTGGGCAACTTTTGGAAAGTTT	1547

```

RESULT 8
US-10-109-616-1
; Sequence 1, Application US/10109616
; Publication No. US20030167484A1
; GENERAL INFORMATION:
; .. APPLICANT: Allen, Keith D.
; .. TITLE OF INVENTION: TRANSGENIC MICE CONTAINING CHANNEL
; .. FILE REFERENCE: R-490
; .. TITLE OF INVENTION: ACTIVATING PROTEASE 1 (CAP1) GENE DISRUPTIONS
; .. CURRENT APPLICATION NUMBER: US/10/109,616
; .. CURRENT FILING DATE: 2002-03-28
; .. PRIOR APPLICATION NUMBER: US 60/280,509
; .. PRIOR FILING DATE: 2001-03-29
; .. PRIOR APPLICATION NUMBER: US 60/311,055
; .. PRIOR FILING DATE: 2001-08-08
; .. NUMBER OF SEQ ID NOS: 4
; .. SOFTWARE: FastSeq for Windows Version 4.0
; .. SEQ ID NO 1
; .. LENGTH: 1797
; .. TYPE: DNA
; .. ORGANISM: Mus musculus
US-10-109-616-1

Query Match      44.6%; Score 744.2; DB 13; Length 1797;
Best Local Similarity 78.4%; Pred.No. 3.2e-191;
Matches 901; Conservative 2; Mismatches 245; Indels 1; Gaps 1;

QY 71 GACACTGCTGCCCTTCCAGCCGGATTCTGGGATCTCTCCCTTGAGCCAACTCTG 130
   |||
Db  8 GAAGCTGCTTCACTTCCAGCCTTAAGTCTGGGATCTCTCACTCTGATCCACACTA 67

QY 131 GGTCTTGCTTGAACACCAACCCCAAGGCTTCTAAGCTTGCCTGGAGTCTGCCAG 190
   |||
Db  68 GATCTGCTTCAAAACACAGCTTCTGTGATCTTCCCTTCACTAAGGAGCTCTTCCGG 127

QY 191 GGGCCCTTGTCTGGGCGATGCGCCMAAGAGGGGATCTCTGGGGGCTCTGGGCACTGGGGCT 250
   |||
Db  128 GGGCCCTTGTCTTAGGCCATGCGCCCAAGGGGTGGGCTCTGGGACTTGGGCACTGGAACT 187

QY 251 GTGGCACTTCTGCTCTAATCTTGGAATTAATCCGCTCGGGGACAGGAGCGGAGCGGACAA 310
   |||
Db  188 GTGACCAATTCGTCTCTTCTTGATTTGCTCCAGTCCGGGAATCCGAGCTGAGAGGAGTGA 247

QY 311 GCTVCTGCGGTGTGGCCCCCAAGACGCAATCAAGGTGAGAGAGTGCAGTGCCTGCT 370
   |||
Db  248 GCTCTGCTGTGGTGGCTGATCAACGCCACGCAATCAACGGTGTGGCAGTGCAGAACCCGCT 307

QY 371 CAGTGGCCCTGGCAGTGCAGATCACTATGAAAGGCGTCAATGTGTGTGGGCTCTTC 430
   |||
Db  308 CAGTGGCCCTGGCAGTGCAGATCACTATGAAAGGCGTCAATGTGTGTGGGCTCTTC 367

```

```

QY 431 GTGTCTGAGCAGTGGGTGTCTGTGAGTGTCTCACTGCTTCCCGAGCAGCAACAGAA 490
DB 368 GTGTCAAAATAATGAGGTGTGTCTGTCTGTCTCACTGCTTCCCGAGAAACAGCAGAGAA 427
QY 491 GCTTATGAGTCAAGCTGGGGGGCCACAGCTTGAATCTCTTCCAGAGACCGCAAGCTC 550
DB 428 GCTTATGAGTCAAGCTGGGGGGCCACAGCTTGAATCTCTTCCAGAGACCGCAAGCTC 487
QY 551 AGCAGCTGAGAGCAATCAATCCCGAGCTTCCAGAGAGGCTTCCAGAGGCTC 610
DB 488 CACAGAGTGTCTAGATCAATCAATCCCGAGCTTCCAGAGAGGCTTCCAGAGGCTC 547
QY 611 ATTGCACTCTCCCACTCAAGCAGAGACCATCACTTCTCCGCTCAATCCGCTCACTG 670
DB 548 ATGCGCTTATCGGCTCAAGCAGAGCTTCTGTCACTTCTCCGCTCAATCCGCTCACTG 607
QY 671 CTCTCTGAGCAGCAGCCTCTCTTCCCGAGGCTTCCAGAGGCTTCCAGAGGCTC 730
DB 608 CTCTCTGAGCAGCAGCCTCTCTTCCCGAGGCTTCCAGAGGCTTCCAGAGGCTC 667
QY 731 CATGTGAGCCCTCCAGTGAAGCTTCTGAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 790
DB 668 CATGTGAGCTTCTCACTGAGCTTCCAGAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 727
QY 791 CTGATCAGTGTGAGAGCTGTGTAATCTGCTGTCAACATCTGAGCAGCAGCAGCAGCAG 850
DB 728 CTCATCAGCAGCAGAACT-GTAGCTGCTGTCAACATTAATGCGGTGCTGAGAAACC 786
QY 851 GCACTTTTTCAGAGAGCACTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 910
DB 787 GCACACTATTCAGAGAGCACTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 846
QY 911 CCAGGAGTACTCTGGGGGGCCACTCTCTGCTGTGAGAGGCTGTGTGTGTGTGTGTGTGT 970
DB 847 CAGAGTGTACTCTGGGGGGCCACTCTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 906
QY 971 CATGTGAGCTGTGAGAGATGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1030
DB 907 CATGTGAGTGTGAGAGATGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 966
QY 1031 CTCAGCTTATGCTCTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1090
DB 967 TCTTACTTATGTCTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1026
QY 1091 CCAGAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1150
DB 1027 CCAGAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1086
QY 1151 TSCCCAGCAGCAGGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1210
DB 1087 AGCGGAGGCTCGGAATTTTGAAGGCTTCTTCTGTGTGTGTGTGTGTGTGTGTGTGT 1146
QY 1211 CCTCTCTC 1219
DB 1147 CCTCTGTG 1155

```

## RESULT 9

```

US-09-922-217-931/c
; Sequence 931, Application US/09922217
; Patent No. US200207614A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Lodee, Michael J.
; APPLICANT: Secrist, Heather
; APPLICANT: Benson, Darin R.
; APPLICANT: Meagher, Madeleine Joy
; APPLICANT: Stolk, John A.
; APPLICANT: Wang, Tonglong
; APPLICANT: Jiang, Yujun
; APPLICANT: Smith, Carole Lynn
; APPLICANT: King, Gordon E.

```

```

; APPLICANT: Wang, Aijun
; APPLICANT: Clapper, Jonathan D.
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND DIAGNOSIS
; TITLE OF INVENTION: OF COLON CANCER AND METHODS FOR THEIR USE
; FILE REFERENCE: 210121.471C13
; CURRENT APPLICATION NUMBER: US/09/922,217
; NUMBER OF SEQ ID NOS: 1124
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO: 931
; LENGTH: 596
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 538
; OTHER INFORMATION: n = A,T,C or G
US-09-922-217-931

```

```

Query Match 35.6%; Score 594.2; DB 9; Length 596;
Best Local Similarity 99.5%; Pred. No. 9,7e-151;
Matches 593; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

```

```

QY 184 GCCCAGAGGGCCCTTGTCTGTGGCCATGAGCAGAGAGGGGCTTGGGGCTTGGCAGCT 243
DB 596 GCCCAGAGGGCCCTTGTCTGTGGCCATGAGCAGAGAGGGGCTTGGGGCTTGGCAGANT 537
QY 244 GGGGGCTGTGGCATTCTGTCTTATCTTGTGATTAATCTCCGCTGGGGAGCAGAGCCGAG 303
DB 536 GGGGGCTGTGGCATTCTGTCTTATCTTGTGATTAATCTCCGCTGGGGAGCAGAGCCGAG 477
QY 304 GGCAGAGCTTCTCGGTGTGGGGCCCGCAGAGCAGCATCAAGAGTGGCAGAGTGTGAGT 363
DB 476 GGCAGAGCTTCTCGGTGTGGGGCCCGCAGAGCAGCATCAAGAGTGGCAGAGTGTGAGT 417
QY 364 GCGCGGTGAGTGGGGCTTGTGAGTGGGGCCCGCAGAGCAGCATCAAGAGTGGTGTGTGT 423
DB 416 CCGCGGTGAGTGGGGCTTGTGAGTGGGGCCCGCAGAGCAGCATCAAGAGTGGTGTGTGT 357
QY 424 CTCTCTGTGTGTGAGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 483
DB 356 CTCTCTGTGTGTGAGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 297
QY 484 CAAGAGCTTATGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 543
DB 296 CAAGAGCTTATGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 237
QY 544 CAAGGTGAGCAGCTTGAAGCAGTATCCCGCAGCTTCTTCCAGAGAGGCTTCCCA 603
DB 236 CAAGGTGAGCAGCTTGAAGCAGTATCCCGCAGCTTCTTCCAGAGAGGCTTCCCA 177
QY 604 GGGGCACTGTGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 663
DB 176 GGGGCACTGTGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 117
QY 664 CATGTGCTCTCCCTGTGAGCAGAGCTTCTTCCCGCAGAGGCTTCCAGAGTGTGTGTGT 723
DB 116 CATGTGCTCTCCCTGTGAGCAGAGCTTCTTCCCGCAGAGGCTTCCAGAGTGTGTGTGT 57
QY 724 CTGGGGTCAATGTGGCCCTTCAAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 779
DB 56 CTGGGGTCAATGTGGCCCTTCAAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1

```

## RESULT 10

```

US-09-833-263-931/c
; Sequence 931, Application US/09833263
; Patent No. US20020110547A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Aijun
; APPLICANT: Clapper, Jonathan D.
; APPLICANT: Stolk, John A.
; APPLICANT: Meagher, Madeleine J.

```

/ TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND  
 / FILE OF INVENTION: DIAGNOSIS OF COLON CANCER AND METHODS FOR THEIR USE  
 / FILE REFERENCE: 210121.471C12  
 / CURRENT APPLICATION NUMBER: US/09/833.263  
 / NUMBER OF SEQ ID NOS: 1093  
 / SOFTWARE: FastSeq for Windows Version 3.0  
 / SEQ ID NO 931  
 / LENGTH: 596  
 / TYPE: DNA  
 / ORGANISM: Homo sapien  
 / FEATURE:  
 / NAME/KEY: misc\_feature  
 / LOCATION: (1)...(596)  
 / OTHER INFORMATION: n = A,T,C or G  
 US-09-833-263-931

Query Match 35.6%; Score 594.2; DB 10; Length 596;  
 Best Local Similarity 99.5%; Pred. No. 9,7e-151;  
 Matches 593; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 184 GCCCCAGGGGCGCTTGTCTGCGGCGCATGCGCAGAGGGGCTCTGGGGCTGGGCGAGCT 243  
 DB 596 GCCCCAGGGGCGCTTGTCTGCGGCGCATGCGCAGAGGGGCTCTGGGGCTGGGCGAGNT 537  
 QY 244 GGGGGCTGTGGCCATTCTGCTTATCTTGATTACTCGGTCGGGAGAGAGCGGAGAG 303  
 DB 536 GGGGGCTGTGGCCATTCTGCTTATCTTGATTACTCGGTCGGGAGAGAGCGGAGAG 477  
 QY 304 GGCAGAAAGCTTCCTGGGTGGCCCCCAAGCAGCATACAGGTGGCAGCTGAGT 363  
 DB 476 GGCAGAAAGCTTCCTGGGTGGCCCCCAAGCAGCATACAGGTGGCAGCTGAGT 417  
 QY 364 CGCGGCTGAGTGGCGCTGGAGGTCAGATCACTATGAGGGGTCCATGTGTGTGG 423  
 DB 416 CGCGGCTGAGTGGCGCTGGAGGTCAGATCACTATGAGGGGTCCATGTGTGTGG 357  
 QY 424 CTCTCTGTGTGAGCAGTGGGTGTGTGAGCTGCTCACTGCTTCCCGAGGAGCACA 483  
 DB 356 CTCTCTGTGTGAGCAGTGGGTGTGTGAGCTGCTCACTGCTTCCCGAGGAGCACA 297  
 QY 484 CAAAGAAAGCTTATGAGGTCAAGCTGGGGGCCCAACAGCTAGACTCTTACCTCCAGAGCGC 543  
 DB 296 CAAAGAAAGCTTATGAGGTCAAGCTGGGGGCCCAACAGCTAGACTCTTACCTCCAGAGCGC 237  
 QY 544 CAAGGTACAGACCTTGAAAGACATCACTCCCAAGCTACCTCCAGAGGGCTCCCA 603  
 DB 236 CAAGGTACAGACCTTGAAAGACATCACTCCCAAGCTACCTCCAGAGGGCTCCCA 177  
 QY 604 GGGCGAATTGCACTCTCCCAACTCAGACAGACCCATCACTTCCCGCTACATCCGAGC 663  
 DB 176 GGGCGAATTGCACTCTCCCAACTCAGACAGACCCATCACTTCCCGCTACATCCGAGC 117  
 QY 664 CATCTGCTCTCCCTGAGCAGCAAGCGCTCTTCCCAAGCGGCTCCACTGCACTGTCACTGG 723  
 DB 116 CATCTGCTCTCCCTGAGCAGCAAGCGCTCTTCCCAAGCGGCTCCACTGCACTGTCACTGG 57  
 QY 724 CTGGGGTCATGTGGCCCCCTCAGTGAAGCTCTGAGCCCAAGCACTGAGAGCAAC 779  
 DB 56 CTGGGGTCATGTGGCCCCCTCAGTGAAGCTCTGAGCCCAAGCACTGAGAGCAAC 1

RESULT 11  
 US-10-025-380-931/C  
 / Sequence 931, Application US/10025380  
 / Publication No. US20020182191A1  
 / GENERAL INFORMATION:  
 / APPLICANT: Xu, Jiangchun  
 / APPLICANT: Lodes, Michael J.  
 / APPLICANT: Benson, Darin R.  
 / APPLICANT: Secrist, Heather  
 / APPLICANT: Meagher, Madeleine Joy  
 / APPLICANT: Stolk, John A.

/ APPLICANT: Wang, Tonglong  
 / APPLICANT: Jiang, Yugu  
 / APPLICANT: Smith, Carole L.  
 / APPLICANT: King, Gordon E.  
 / APPLICANT: Wang, Aijun  
 / APPLICANT: Clapper, Jonathan D.  
 / APPLICANT: Skelky, Yasir A. W.  
 / APPLICANT: Fanger, Gary R.  
 / APPLICANT: Vedvick Thomas S.  
 / APPLICANT: Carter, Darrick  
 / TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND DIAGNOSIS  
 / FILE OF INVENTION: OF COLON CANCER AND METHODS FOR THEIR USE  
 / FILE REFERENCE: 210121.471C14  
 / CURRENT APPLICATION NUMBER: US/10/025,380  
 / CURRENT FILING DATE: 2001-12-19  
 / NUMBER OF SEQ ID NOS: 1129  
 / SOFTWARE: FastSeq for Windows Version 4.0  
 / SEQ ID NO 931  
 / LENGTH: 596  
 / TYPE: DNA  
 / ORGANISM: Homo sapiens  
 / FEATURE:  
 / NAME/KEY: misc\_feature  
 / LOCATION: 538  
 / OTHER INFORMATION: n = A,T,C or G  
 US-10-025-380-931

Query Match 35.6%; Score 594.2; DB 14; Length 596;  
 Best Local Similarity 99.5%; Pred. No. 9,7e-151;  
 Matches 593; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 184 GCCCCAGGGGCGCTTGTCTGCGGCGCATGCGCAGAGGGGCTCTGGGGCTGGGCGAGCT 243  
 DB 596 GCCCCAGGGGCGCTTGTCTGCGGCGCATGCGCAGAGGGGCTCTGGGGCTGGGCGAGNT 537  
 QY 244 GGGGGCTGTGGCCATTCTGCTTATCTTGATTACTCGGTCGGGAGAGAGCGGAGAG 303  
 DB 536 GGGGGCTGTGGCCATTCTGCTTATCTTGATTACTCGGTCGGGAGAGAGCGGAGAG 477  
 QY 304 GGCAGAAAGCTTCCTGGGTGGCCCCCAAGCAGCATACAGGTGGCAGCTGAGT 363  
 DB 476 GGCAGAAAGCTTCCTGGGTGGCCCCCAAGCAGCATACAGGTGGCAGCTGAGT 417  
 QY 484 CAAAGAAAGCTTATGAGGTCAAGCTGGGGGCCCAACAGCTAGACTCTTACCTCCAGAGCGC 543  
 DB 296 CAAAGAAAGCTTATGAGGTCAAGCTGGGGGCCCAACAGCTAGACTCTTACCTCCAGAGCGC 237  
 QY 544 CAAGGTACAGACCTTGAAAGACATCACTCCCAAGCTACCTCCAGAGGGCTCCCA 603  
 DB 236 CAAGGTACAGACCTTGAAAGACATCACTCCCAAGCTACCTCCAGAGGGCTCCCA 177  
 QY 604 GGGCGAATTGCACTCTCCCAACTCAGACAGACCCATCACTTCCCGCTACATCCGAGC 663  
 DB 176 GGGCGAATTGCACTCTCCCAACTCAGACAGACCCATCACTTCCCGCTACATCCGAGC 117  
 QY 664 CATCTGCTCTCCCTGAGCAGCAAGCGCTCTTCCCAAGCGGCTCCACTGCACTGTCACTGG 723  
 DB 116 CATCTGCTCTCCCTGAGCAGCAAGCGCTCTTCCCAAGCGGCTCCACTGCACTGTCACTGG 57  
 QY 724 CTGGGGTCATGTGGCCCCCTCAGTGAAGCTCTGAGCCCAAGCACTGAGAGCAAC 779  
 DB 56 CTGGGGTCATGTGGCCCCCTCAGTGAAGCTCTGAGCCCAAGCACTGAGAGCAAC 1

RESULT 12



```
; Sequence 7606, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR G
; FILE REFERENCE: A60MCA-X-2
; CURRENT APPLICATION NUMBER: US/10/029,386
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 34288
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 7606
; LENGTH: 543
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO CHR16.3
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 5.1
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 5.7
; OTHER INFORMATION: EXPRESSED IN PETAL LIVER, SIGNAL = 4.3
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 4.9
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 3.3
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 3.6
; OTHER INFORMATION: SWISSPROT HIT: Q16551, EVALUE 7.00e-04
; OTHER INFORMATION: EST HUMAN HIT: AU142128.1, EVALUE 0.00e+00
; OTHER INFORMATION: NT HIT: U33446.1, EVALUE 0.00e+00
US-10-029-386-7606
```

```
Query Match 17.7%; Score 294.6; DB 13; Length 543;
Best Local Similarity 98.0%; Pred. No. 1e-69;
Matches 297; Conservative 1; Mismatches 5; Indels 0; Gaps 0;
```

```
QY 1 CACACTGCTGCTTGGATTAATCTCAGGCGTCTCCGTTGGGCGCGCTCCCTGCTTGAAG 60
DB 417 CACACTGCTGCTTGGATTAATCTCAGGCGTCTCCGTTGGGCGCGCTCCCTGCTTGAAG 358
QY 61 GCCAGCCTTGAGACCTTGCGCCCTTCCAGCCCGGATTCCTGGGATCTTCCCTCTGAG 120
DB 357 GCCAGCCTTGAGACCTTGCGCCCTTCCAGCCCGGATTCCTGGGATCTTCCCTCTGAG 298
QY 121 CCAACATCTGGGTCCTGCTTGCACACACCCCAAGGCTTCTAATCTGCGTGCCTGAG 180
DB 297 CCAACATCTGGGTCCTGCTTGCACACACCCCAAGGCTTCTAATCTGCGTGCCTGAG 238
QY 181 TGTGCCCCAGGGGCGCTTGTCTTGCGCCATGCGCMAAGAGGGGCTCTGGGGCTTGAGCA 240
DB 237 TGTGCCCCAGGGGCGCTTGTCTTGCGCCATGCGCMAAGAGGGGCTCTGGGGCTTGAGCA 178
QY 241 GCTGGGGGCGTGTGGCCATTCCTAATCTGATTAATCTCCGCTCGGGGACAGAGCGGA 300
DB 177 GCTGGGGGCGTGTGGCCATTCCTAATCTGATTAATCTCCGCTCGGGGACAGAGCGGA 118
QY 301 AGG 303
DB 117 AGG 115
```

```
RESULT 15
US-10-029-386-21306/c
; Sequence 21306, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR G
; FILE REFERENCE: A60MCA-X-2
; CURRENT APPLICATION NUMBER: US/10/029,386
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 34288
```

```
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 21306
; LENGTH: 315
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO CHR16.3
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 5.1
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 5.7
; OTHER INFORMATION: EXPRESSED IN PETAL LIVER, SIGNAL = 4.3
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 4.9
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 3.3
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 3.6
; OTHER INFORMATION: NT HIT: g114779472, EVALUE 0.00e+00
; OTHER INFORMATION: SWISSPROT HIT: Q16551, EVALUE 1.00e-08
; OTHER INFORMATION: EST_HUMAN HIT: AU142128.1, EVALUE 0.00e+00
US-10-029-386-21306
```

```
Query Match 17.6%; Score 293.6; DB 13; Length 315;
Best Local Similarity 99.7%; Pred. No. 1.7e-69;
Matches 293; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 CACACTGCTGCTTGGATTAATCTCAGGCGTCTCCGTTGGGCGCGCTCCCTGCTTGAAG 60
DB 294 CACACTGCTGCTTGGATTAATCTCAGGCGTCTCCGTTGGGCGCGCTCCCTGCTTGAAG 235
QY 61 GCCAGCCTTGAGACCTTGCGCCCTTCCAGCCCGGATTCCTGGGATCTTCCCTCTGAG 120
DB 234 GCCAGCCTTGAGACCTTGCGCCCTTCCAGCCCGGATTCCTGGGATCTTCCCTCTGAG 175
QY 121 CCAACATCTGGGTCCTGCTTGCACACACCCCAAGGCTTCTAATCTGCGTGCCTGAG 180
DB 174 CCAACATCTGGGTCCTGCTTGCACACACCCCAAGGCTTCTAATCTGCGTGCCTGAG 115
QY 181 TGTGCCCCAGGGGCGCTTGTCTTGCGCCATGCGCMAAGAGGGGCTCTGGGGCTTGAGCA 240
DB 114 TGTGCCCCAGGGGCGCTTGTCTTGCGCCATGCGCMAAGAGGGGCTCTGGGGCTTGAGCA 55
QY 241 GCTGGGGGCGTGTGGCCATTCCTAATCTGATTAATCTCCGCTCGGGGACAGAGCGGA 294
DB 54 GCTGGGGGCGTGTGGCCATTCCTAATCTGATTAATCTCCGCTCGGGGACAGAGCGGA 1
```

Search completed: December 16, 2003, 01:14:51  
Job time : 555 secs